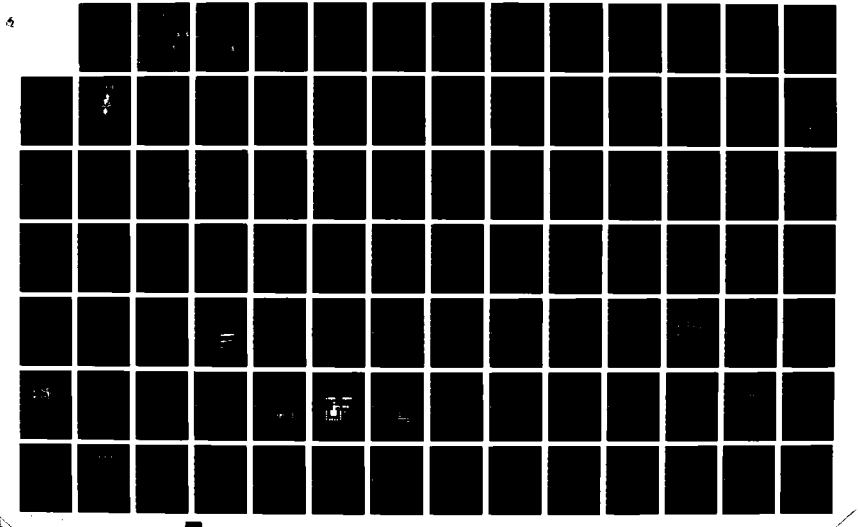
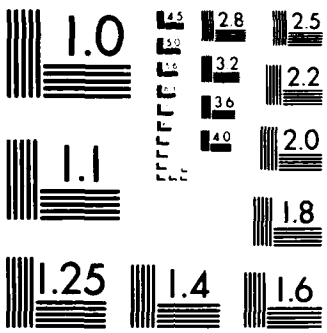
AD-A188 871 A THREE DIMENSIONAL ELECTRONIC RETINA ARCHITECTURE(U) 1/1
AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH
G L SALADA DEC 87 AFIT/GCS/ENG/87D-23

UNCLASSIFIED F/G 12/9 NL





A THREE DIMENSIONAL ELECTRONIC RETINA ARCHITECTURE

Gray L. Salada
Captain, USA
AFIT/GCS/ENG/87D-23

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GCS/ENG/87D-23

A THREE DIMENSIONAL

ELECTRONIC RETINA ARCHITECTURE

Gray L. Salada
Captain, USA
AFIT/GCS/ENG/87D-23

Approved for public release; distribution unlimited

AFIT/GCS/ENG/87D-23

A THREE DIMENSIONAL ELECTRONIC RETINA ARCHITECTURE

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Electrical Engineering

Gray L. Salada, B.S.

Captain, USA

December 1987

TABLE OF CONTENTS (continued)

ACKNOWLEDGEMENTS

This research effort could not have been completed without the help and understanding of several people. I would first like to thank my loving wife Pamela for all her support and patience.

I am grateful to the other members of the VLSI group: Captain Keith Jones, Captain Dave Gallagher, Second Lieutenant Scott Hauser, Captain Craig Spanburg, and Captain Larry Shand for their assistance and encouragement.

A special thanks to my thesis advisor Captain Richard Linderman for his guidance and encouragement. I would also like to acknowledge the the members of my thesis committee Doctor Matthew Kabrisky, and Captain Steve Rogers. Finally, I would like to thank Massimo Sivilotti at CALTECH for his explanation of the work accomplished there.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

ABSTRACT

The biological retina is a compact, real-time image processor. It performs analog computations in parallel in depth. Such processing capabilities have so far been unattainable in comparably sized man-made image processors. This thesis investigates using VLSI technology and a three dimensional architectural approach in designing an electronic retina.

This effort involves the comparison of alternative 3D designs and the implementation of VLSI circuitry capable of low-level image processing. The 3D architectures compared include a model based on the Hughes 3D computer, a model using a monolithic wafer with multiple circuit layers, and a model using stacked wafers with fiber optic interconnects. A conventional 2D architecture is presented for comparison/reference.

Circuits for photo detection, and spatial and temporal processing were implemented. It was shown that using VLSI analog circuits and a 3D architecture it is feasible to fabricate a real-time image processor in a compact package.

A THREE DIMENSIONAL ELECTRONIC RETINA ARCHITECTURE

CHAPTER 1

Introduction

1.1 Background

The operation of the human eye is a functional marvel. The visual system has the highest capacity for parallel processing of all the senses. It consists of a network of hundreds of successive two dimensional arrays each consisting of millions of interconnected parallel computers. As the actual visual processing is transparent to healthy humans, sight seems almost instantaneous [Mon82].

The retina where much of this processing takes place is the ultimate low-level image processor. Figure 1 shows the different layers and the major components of the retina. In the human retina, the receptors (where the image is input) consist of 125,000,000 rods and cones. When the retina finishes its processing at the ganglion cells, there are only 1,000,000 outputs to the brain. This represents a data reduction of 125 to 1. Only that data which carries the essence of the image is forwarded to the brain on the optic-nerve fibers [Kab66].

Considerable research effort has been expended, and is being expended, on designing systems which emulate functions of the human visual system. While the retina performs

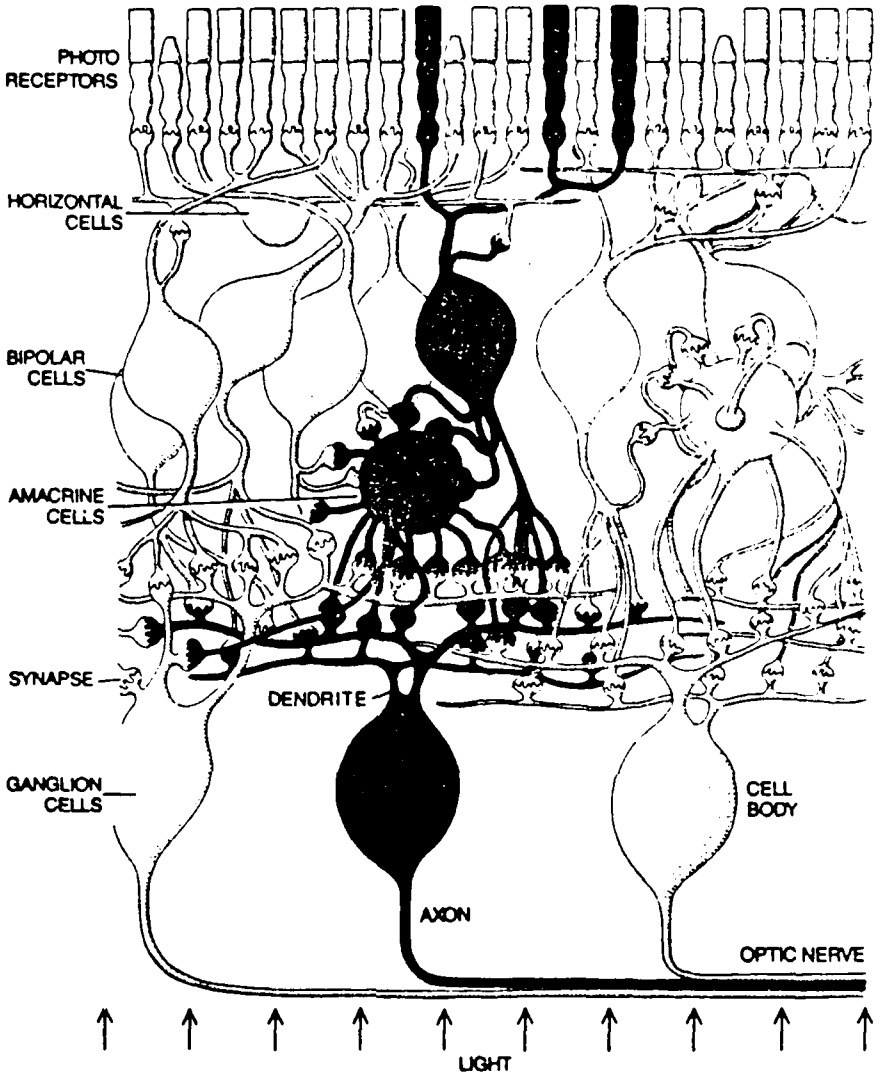


Figure 1-1: RETINA ARCHITECTURE
(from [Pog87])

low level processing and data compression of images seemingly instantly, digital computers cannot match this performance on comparable scenes. Furthermore, the space occupied by the equipment required to do even minimal processing cannot fit in an area as compact as the human skull. An example of the necessary digital computational power presently required for a trivial (as far as the human visual system is concerned) image

processing problem is:

> A 512 pixel by 512 pixel input image delivered at 30 frames per second with 256 levels per pixel implies a data rate of 7.86 million bytes per second. Typical low-level IU (image understanding) algorithms currently in use require 100 to 500 operations on each pixel for a total throughput of 786 million to 3.93 billion operations per second (BOPS)! .. These demands have deterred the real time implementation of even the most simple IU algorithms [Weh85].

The example illustrates the need for efficient algorithms and architectures with higher throughput if the retina is to be emulated. Advances in Very Large Scale Integration (VLSI) and also 3D computer architectures have brought about new possibilities for designing an electronic retina. A novel architecture which combines the power of VLSI with the high throughput of the 3D computer is the three dimensional electronic retina. The 3D electronic retina is a cellular architecture which stacks custom VLSI signal processing wafers directly behind the focal plane array (FPA) in several layers (see Figure 1-2). The cellular electronic retina provides a massively parallel architecture, quite similar to the eye's general architecture in which processing is conducted in parallel in depth. Furthermore, as with the eye, the 3D electronic retina processes the data smoothly and continuously using analog circuitry.

1.2 Problem Statement

The problem is to design within the constraints of VLSI technology, a 3D electronic retina capable of performing rudimentary preprocessing of multi-dimensional data providing spatial and temporal image information as well as data compression of the scene.

1.3 Current Literature

While there has been considerable effort in investigating the structure and functions of the biological eye; most of the effort in image processing has concentrated on using digital devices and processing techniques as opposed to designing structures which

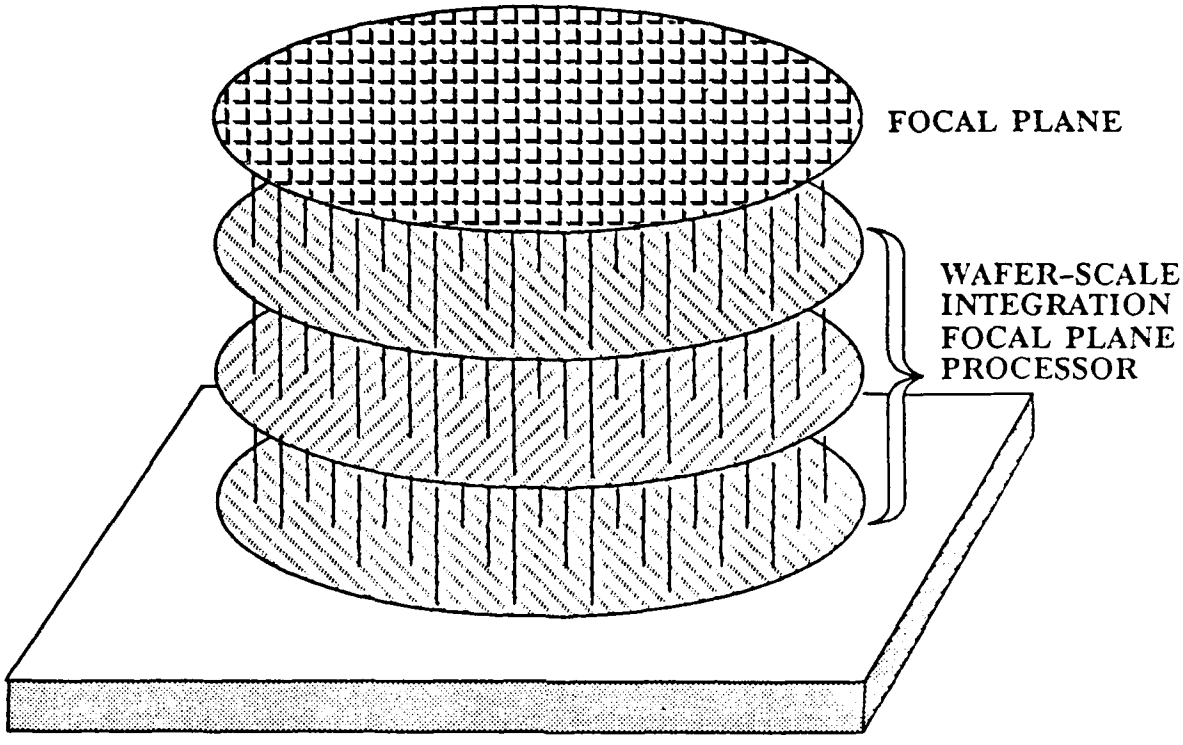


Figure 1-2: STACKED WAFER CONCEPT
(from [Deg86])

evaluate the eye's functions in an analogous manner. As alluded to previously, this may have been due more to technical feasibility than desirability. In any case, while it has yet to be proven which approach is the better, it is likely that there are applications which are more suitable for an actual electronic eye versus a digital simulation. This literature review uncovered research which has been done in more areas essential to the electronic retina including work done in actual retina processing elements, 3D architecture, and photosensors.

1.3.1 Electronic Retina. There have been no 3D electronic retinas designed and implemented. However, Carver Mead and students at CALTECH designed and implemented a 2D electronic retina. This design uses devices fabricated in silicon using CMOS. Retina elements emulated in the CALTECH design are receptor cells (cones), horizontal cells, and amacrine cells. Though not functionally complete, this rudimentary retina is one of the first attempts at a silicon electronic retina [Mea87].

The key approach taken in Mead's design is the use of analog circuits to process the image data. This more closely approximates how the biological eye processes data, and represents a departure from typical image processing techniques. His work on the electronic retina builds on work Richard Lyon did, which used the biological concept of lateral inhibition. Lyon designed an optical mouse as an alternative to the electromechanical mouse which was available at the time. The inhibition is implemented using digital logic and performs a local area gain control based on the inhibition radius. Figure 1-3 shows a two pixel imager with mutual inhibition. The inhibition radius determines how may pixels are in a neighborhood. For example, a pixel with an inhibition radius that included its eight nearest neighbors would use 9 input nor gates [Lyo81].

Another technique for fabricating an electronic retina involves charge-coupled computing. Recent work at Columbia University has produced analog computers which use charge-coupled devices (CCD's) to perform arithmetic and logic functions. By designing, at 1.5 micron design rules, a FPA with a charge-coupled computer per pixel element it may be possible to obtain real time image processing at high performance frame rates (4000 Hz) with adequate resolution for many applications. Higher resolutions may be possible at human vision rates of 100 Hz by assigning more pixel elements per charge-coupled computer, thus increasing the photo sensitive area [Fos87].

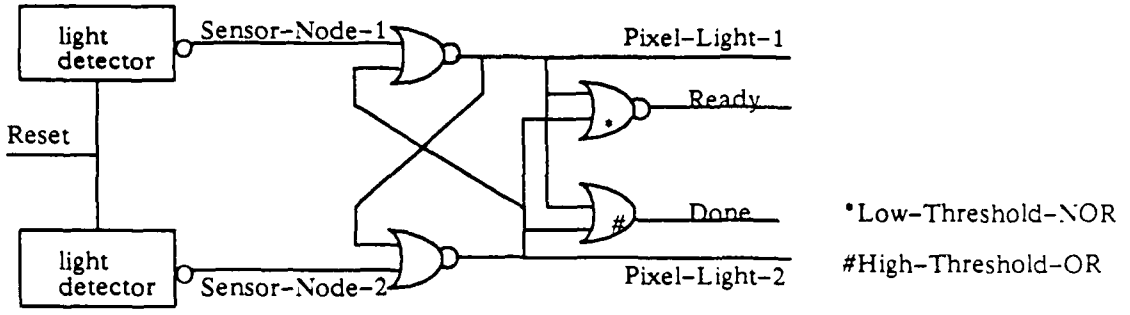


Figure 1-3: TWO PIXEL DIGITAL INHIBITION
(from [Lyo81])

1.3.2 3D Device Technology. A major stumbling block in designing a 3D electronic retina is the third dimension. Until recently, it was not technically feasible to fabricate 3D devices (devices which have circuitry in the x, y, and z planes). Kataoka and others at the Electrotechnical Laboratory in Japan are investigating a 3D architecture for an artificial retina [Kat85].

The 3D architecture they are exploring is a layered silicon on insulator (SOI) design. In this process, layers of VLSI devices are fabricated on an insulator layer placed on top of the previous finished layer of devices. Connections between the two layers are made as necessary. This process is continued until all the layers are fabricated. The top layer consists of the photodiodes [Kat85]. Figure 1-4 shows the general steps in the process of fabricating a layer of devices on top of another layer of devices, and then connecting them along the z-axis. This type of 3D architecture has very promising possibilities for imaging devices with 3 to 5 active layers but, is very much in the research stage and is

not expected to be feasible until the mid to late 1990's [Gri84]. Major limitations involve heat dissipation and through-hole technology.

Work at Hughes Research Laboratories has led to the successful fabrication of a 3D computer using a unique concept which is presently feasible. The work by Jan Grinberg

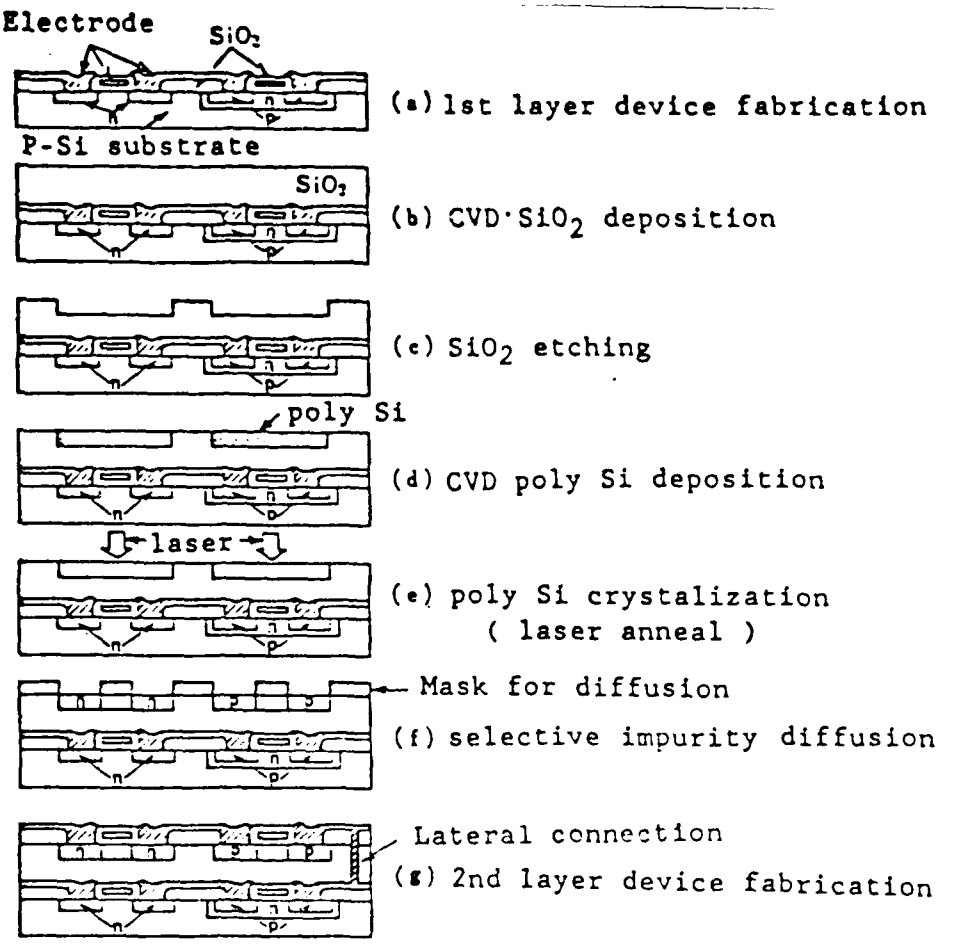


Figure 1-4: 3D SOI DEVICE FABRICATION
(from [Kat85])

and others of Hughes Research Laboratories is the genesis of this research effort. Their design is a Single Instruction Multiple Data (SIMD) computer that assigns one processor per pixel element in the FPA when used for image processing. While there are some systems which presently use this massively parallel concept such as the ILLIAC IV, ILL's DAP machine, UCL's CLIP-4, and Goodyear's MPP; the Hughes design is radically different in that it uses VLSI to implement the processors, and the processors are arrayed vertically in a stack of silicon wafers each of which contains an array of processor elements. This arrangement of processors reduces the average interconnect length from 0(N) to $0(N^{1/3})$ where N is the number of circuit nodes [Gri4A]. Advantages of the three dimensional architecture includes ease of programming, low power consumption, inexpensive construction, and an n^2 speedup over conventional single processor architectures operating on two dimensional data sets [Gri84].

Additionally, the 3D computer offers significant space savings. A conventional computer using a backplane with various circuit boards may have as little as 1% active silicon circuitry per total computer volume. In the 3D computer over 90% of the volume is active silicon [Cor86].

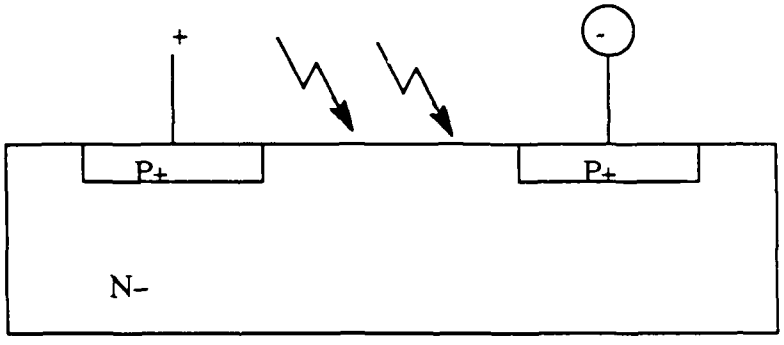
To implement the 3D computer concept, Hughes developed a proprietary interconnection design for the stacked wafers which allows a high bandwidth channel for communications between wafers. The interconnection devices consist of a series of microbridges which allow space between wafers to accommodate potential wafer warp while also cutting parasitic impedance which lowers the propagation delay and power dissipation of the signal drivers [Gri4A]. Appendix A describes the 3D fabrication process and details of implementation.

1.3.3 Focal Plane Technology. Other work that has contributed to the ability to enhance image processing both conventionally and possibly with an electronic retina are advances in imaging arrays such as Charge-Coupled Devices (CCD's) and Charge-Injection Devices (CID's). There is a tremendous amount of literature concerning focal plane arrays, with most recent articles concerning CCD's and CID's. A description of these sensors and their operation is described in the next chapter. However, it should be noted that present focal plane array technology allows the manufacture of sensors sensitive from 1Å to 11,000Å (soft X-ray to near infrared) [Jan87] and arrays containing 3,000,000 (1732 x 1732) detector elements [Key80].

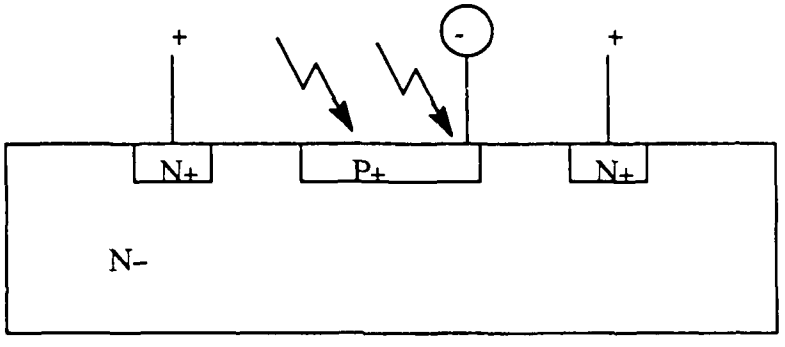
For the older photodiode/phototransistor technology, Bergman and others have outlined three detector types suitable for CMOS VLSI technology. These include the lateral p+n-p+ phototransistor, the vertical p+n- photodiodes, and the lateral p+n-p+ photodiode. Each detector type has advantages and disadvantages which need to be considered in a system design. The lateral phototransistor provides large photocurrent but has slow response time. The lateral photodiode provides high speed but has relatively poor sensitivity. The vertical photodiode lies between the other types in regard to sensitivity and response time [Ber86]. Figure 1-5 shows the three general types.

1.4 Assumptions

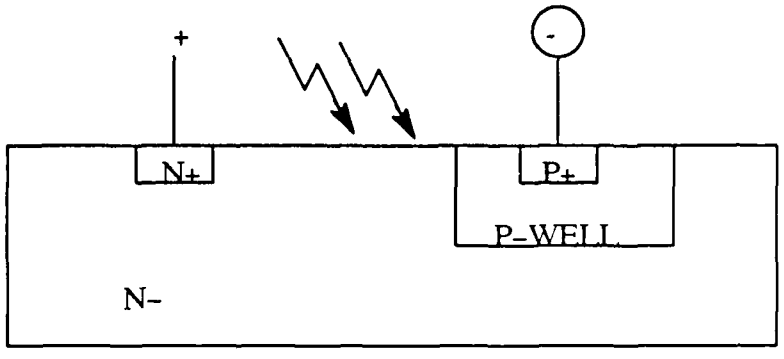
This study assumes a maximum vertical bus data rate of 10 MHz in the three dimensional processor. This constraint is imposed by the impedance restrictions applicable to the z-axis interconnects available with current technology. As this is approximately two orders of magnitude slower than horizontal data rates, it points to the need for efficient partitioning of the architecture in the three available dimensions [Gri84]. Also, the availability of 512 x 512 wafer feedthrough technology in the 3D computer is



Lateral p+n-p+ Phototransistor



Vertical pn Diode



Lateral pn Diode

Figure 1-5: CMOS COMPATIBLE DETECTORS
(from [Ber86])

assumed, as is the availability of optical systems which approximate the retina optics.

1.5 Scope

This thesis considers images delivered by 512 x 512 arrays of pixel elements. This image size provides a representative example which may be extrapolated to other image sizes as technology evolves. Additionally, color images are not addressed directly, however, they may be added to the selected architecture by modifying the preprocessor and adding the appropriate filters in the lens system or the focal plane.

1.6 Approach

An architecture will be designed which facilitates the fabrication of an electronic retina. Using the CALTECH 2D electronic retina cell as an electronic subcomponent, various general architectures will be devised and capability to perform retinal functions examined. The most promising architecture will be selected and presented in greater detail. Emphasis will be given more to a functional capability than to a direct mapping of the biological retina to electronic component. Also, emphasis will be placed on architectures which may be fabricated in the near term technology.

1.7 Materials and Equipment

The design of the VLSI retina architecture will be done using AFIT computer resources and the AFIT VLSI computer aided design (CAD) software environment. All these resources are available.

1.8 Summary

The 3D electronic retina brings together the most advanced technologies in VLSI design and 3D computer architecture. The implementation of the 3D electronic retina not only advances the state of the art in image processing by providing a very fast image

preprocessor, but also advances research in neural networks by providing a new 3D hardware approach to analog problems.

1.9 Sequence of Presentation

This introductory chapter will be followed by Chapter II which describes the theory of how the retina functions, and how it is mapped into an electro-optic design. Chapter III will explain various 3D architecture and Chapter IV will cover the VLSI design considerations. Finally, Chapter V will explore the results and provide recommendations for further research in this area.

CHAPTER 2

Detailed Analysis of the Problem

2.1 Overview

This chapter describes the biological retina anatomy and functions prior to discussing the electro-optic design process. The two are thoroughly blended providing the basis for designing architectures. The elements needed in the electro-optic system are thoroughly presented as they apply to the electronic retina design.

2.2 Biological Retina

As mentioned previously, the retina is an extremely complex instrument consisting of five major cell types (see Figure 1-1) interacting to produce imagery data for the brain to interpret. An understanding of the function of these different cell types and how they interact is a prerequisite for designing an electronic counterpart. Since the design only implements the spatial and temporal aspects of retinal processing, only those functions will be explained in detail. The retina functions which correspond to spatial and temporal processing are excitation and inhibition. This chapter describes the anatomy of the retina and the excitory and inhibitory functions of the retina.

2.2.1 Retina Anatomy. Figure 2-1 shows the location of the retina in the eye, which is essentially the back lining of the eye. The retina is composed of five major cell types: receptor cells, bipolar cells, horizontal cells, amacrine cells, and ganglion cells. Junctions where cells make connections to each other are called synapses. At these synapses the communication between cells takes place by chemical or electrical

mechanisms [Pog87]. The general throughput of the retina is along an axis starting from the receptor cells through the bipolar cells to the ganglion cells, after which the information is transmitted to the brain. A signal traveling along this general path is influenced first by the horizontal cells in the outer synaptic layer (where receptor cell and top bipolar cell synapses are located) and then by the amacrine cells at the inner synaptic layer (where the ganglion cells and the other end of the bipolar cells synapses are located). Note that light actually passes through the other four cell types before reaching the receptor cells.

The receptors are distributed throughout the area of the retina as in Figure 2-2 except for a small area where the optic nerve connection is located. This area is a blind spot in the eye. The rest of the area can be roughly divided into two regions, the fovea

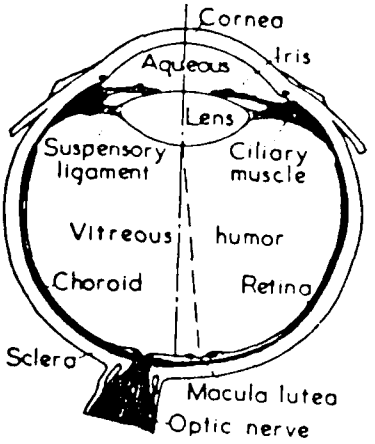


Figure 2-1: LOCATION OF THE RETINA IN THE EYE
(from [Bea82])

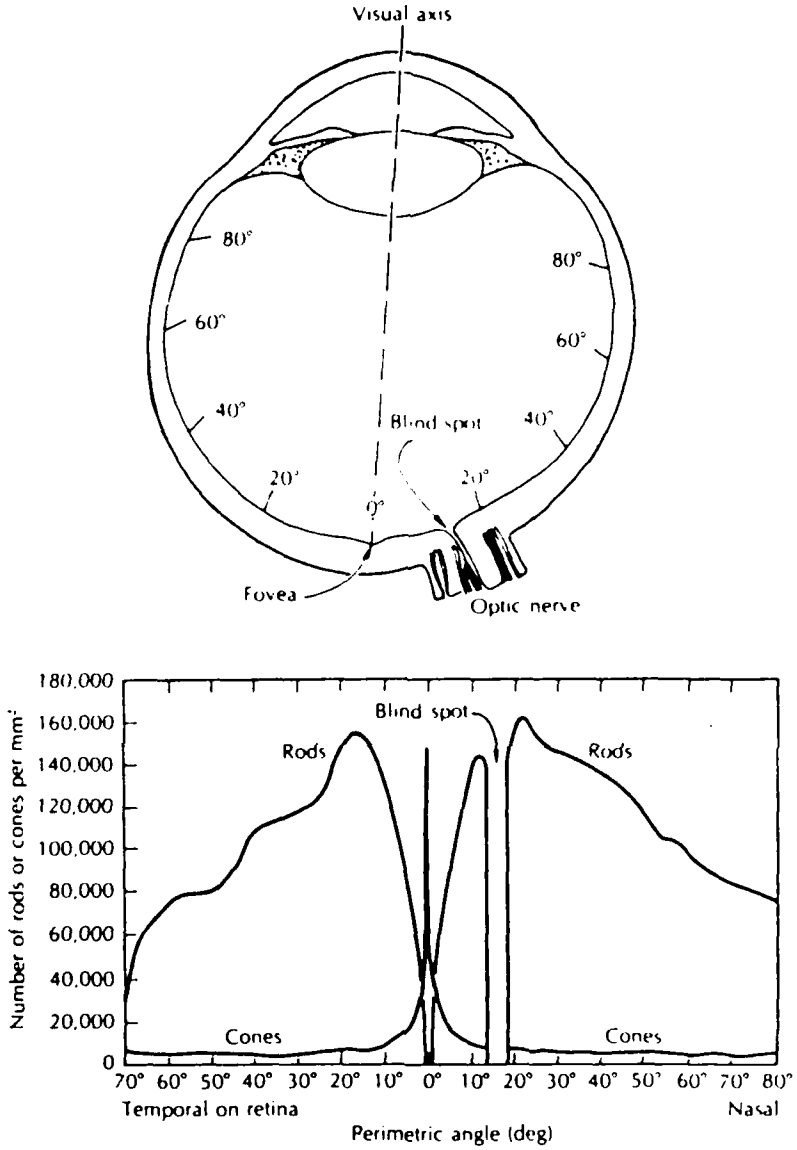


Figure 2-2: DISTRIBUTION OF RODS AND CONES
(from [Cor70])

and the peripheral. The fovea is that area (about .3 mm in diameter) of highest visual acuity and is concerned mainly with form and color data. It has approximately 2° field of view out of the total field of about 180° horizontal and 60° vertical. For comparison

purposes, the image of a full moon is about 0.2 mm on the retina. There are approximately 100,000 receptors (all cones) in the fovea and each is mapped to a corresponding "sole user" channel in the optic nerve. The peripheral consist of the other receptors (approximately 124.9 million) distributed in an array which becomes less dense towards the edge of the field of view. These peripheral receptors are mapped to about 900,000 channels in the optic nerve. This necessitates a tremendous amount of data reduction. The peripheral receptors mainly deal with motion detection in the image area not being focused on the fovea [Kab66].

2.2.1.1 Receptor Cells. The two types of receptor cells are rods and cones. The distribution of the rods and cones in the retina is shown in Figure 2-2. The rods and cones are sensitive to different spectral ranges (Figure 2-3) which overlap and thus provide broader vision capabilities. Overall, normal human vision is generally considered to be sensitive to wavelengths from 390 nm to 780 nm. Rods operate at lower light levels providing "night" vision while cones operate at higher light levels and provide "daylight" vision as well as color and high acuity vision. Rods are about .002 mm in diameter and cones are about .006 mm in diameter except for in the densely packed fovea where they are between .0015 mm and .0030 mm in diameter [Hec87].

2.2.1.2 Bipolar Cells. The bipolar cells are the throughways from the outer synapse layer to the inner synapse layer. They are the only cell type that bridges the two layers [Mas86]. Thus a bipolar connects a receptor field to a ganglion cell. A receptor field is that area consisting of receptor cells mapped to a common ganglion cell.

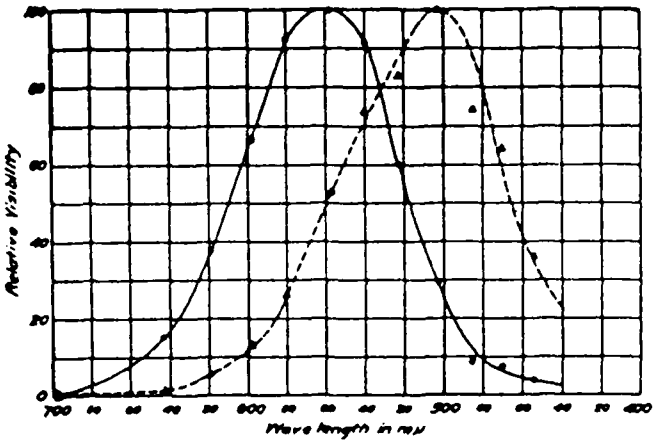


Figure 2-3: SPECTRAL SENSITIVITY OF RETINA
Dashed curve, rods; solid curve, cones; from [Bea82].

2.2.1.3 Horizontal Cells. The horizontal cells interact with receptors and bipolar at their common synapse. As their name implies, they interact over a wide area by spreading horizontally in the outer synapse layer [Wer73]. Their primary purpose is to modify receptor and bipolar signals in proportion to spatial and temporal light intensity in surrounding receptors.

2.2.1.4 Amacrine Cells. Amacrine cells are similar to horizontal cells in that they operate over a wide horizontal area. They interact with bipolars and ganglion cells in the inner synapse layer. There are as many as 30 distinct types of amacrine cells, each having a specialized function. The exact functions of the various types are not entirely known; however, they are believed to be key factors in motion detection [Mas86].

2.2.1.5 Ganglion Cells. The ganglion cells are the interface between the retina and the optic nerve. There are 1 million ganglion cells corresponding to the 1 million channels in the optic nerve. They send a continuous stream of impulses on their assigned channel in the optic nerve. This stream of pulses varies in relation to the intensity of light on those cells acting in its receptive field. There are different types of ganglion cells. Some of the cells simply pass information directly from the bipolar cells into the optic nerve. Other ganglion cells are affected by amacrine cells. These cells respond to movement or changes in the scene [Wer73].

2.2.2 Retina Functions. In general, to look at an object, that object is focused on the fovea. If the object is too large, then the eye will move, scanning the object across the fovea. An experiment which illustrates the operation of the fovea as compared to the peripheral is "for instance, if you concentrate on the letter "n" in the center of the word concentrate then you will be unable to read the letters at the beginning and end" [Kab66]. The letter "n" is focused on the fovea providing full detail of form and color while the "c" and ending "e" are "blurred" because they are imaged on the peripheral. Note, however, if the "c" or "e" were able to move in relation to the rest of the letters, the peripheral would register this information and the "distraction" most likely would result in that area being focused on the fovea.

Another experiment which illustrates the operation of the fovea, as well as the different sensitivities of cones and rods, is to look at a star in the night sky. It is often seen better by looking a little to one side of it so that the image is not focused on the fovea where there are no rods. Recall that the rods are the receptors sensitive to low level light.

Only a small portion of the immense amount of data incident on the receptors makes its way to the brain. This residual data, however, contains all the essential information for the brain to visually analyze the environment - to learn and survive. Exactly how the retina accomplishes its mission is not totally known, although it has been determined mutual inhibition plays a part.

2.2.2.1 Mutual Inhibition. Mutual inhibition is thought to be a method by which the retina filters out the most essential data from the scene. This information primarily concerns contours, which is the essence of any picture [Rat61]. In general, inhibition is the process where receptors effect other receptors in a given area by modifying their response. The inhibition is not only a factor of the spatial distribution of light on the receptor mosaic, but also of temporal changes in the light intensities on the mosaic. The eye of the Limulus (horseshoe crab) because of its large size and simplicity has been used to demonstrate this function [Rat61].

2.2.2.2 Spatial Processing. The horizontal cell is the means by which lateral communication takes place between the essentially vertically oriented receptors and bipolars. If a receptor is activated, it "fires" with a corresponding impulse. If, at the same time, a neighboring receptor is activated, the two receptors' impulses interact by means of the horizontal-bipolar-receptor synapses to reduce the impulse. The amount of interaction is dependent upon the distance between the receptors, as the amount of interaction reduces with distance [Rat61].

The response r_1 of a receptor which fires simultaneous with a receptor r_2 may be determined by the following equation:

$$r_1 = e_1 - K_{1,2}(r_2 - r_{1,2}^0) \tag{2-1}$$

where e_1 is the external stimulus applied to the receptor r_1, $K_{1,2}$ is the coefficient of inhibiting action of receptor 2 on receptor 1, and $r^0_{1,2}$ is the threshold level at which receptor 2 begins to inhibit receptor 1 [Rat61]. $K_{p,j}$ and $r^0_{p,j}$ are the factors in the equation which take the distance between receptors into account; $r^0_{p,j}$ increases with increasing distance and $K_{p,j}$ decreases with increasing distance [Rat61]. The response of r_2 may be determined by appropriate substitution into Eq (2-1) resulting in:

$$r_2 = e_2 - K_{2,1}(r_1 - r^0_{2,1}) \tag{2-2}$$

The inhibition does not occur between just two receptors but is a result of the interaction of many receptors in an area. The response of a receptor may be determined by the sum of the individual inhibiting contributions from all the receptors and is given by the following equation:

$$r_p = e_p - \sum_{j=1}^{n} K_{p,j}(r_j - r^0_{p,j}) \tag{2-3}$$

where p = 1,2,3,...,n; and $p \neq j$ [Rat61].

2.2.2.3 Temporal Processing. In addition to the spatial summation just noted, a temporal response is elicited by changing illumination. Whenever a change occurs, a large transitional response follows it. If the illumination increased, then an exaggerated increased transient response occurs; if the illumination decreased then an exaggerated decreased transient response occurs. These transient responses have an opposite effect on the response of neighboring elements. The transients quickly settle to the steady state level expected by the new levels of illumination [Rat61]. Figure 2-4 shows the temporal response of two neighboring receptors in a Limulus. The solid dots represent the steady state receptor and the open dots represent the changing receptor. Note the sharp up peak resulting from increased illumination for two seconds and then

the down peak resulting from decreased illumination. Also note the opposite effect the transient has on a neighboring receptor under steady state conditions. These peaks are transient and the response quickly settles to a steady state value.

2.3 Electro-optic Design

The functional blocks of an electro-optic design for a vision system are depicted in Figure 2-5. The portions of the system design described in this report are the detectors and the low- level retinal processing electronics. These are the parts that will implement the biological retinal elements discussed above. However, for purposes of understanding the complexity of the final architecture all the functional blocks will be discussed.

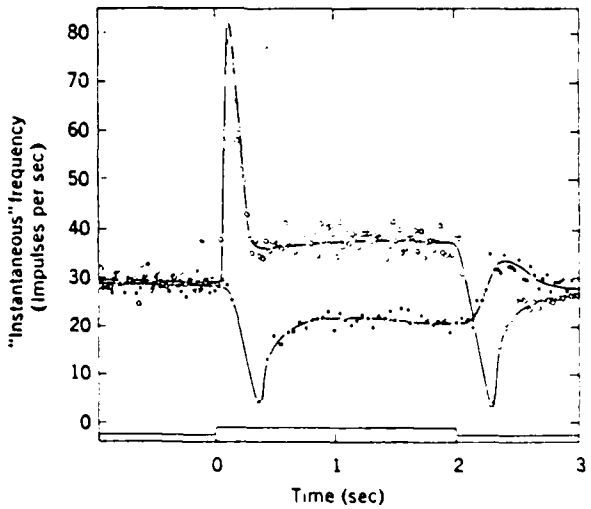


Figure 2-4: TEMPORAL EFFECT IN RETINA
(see text for explanation; from [Rat61])

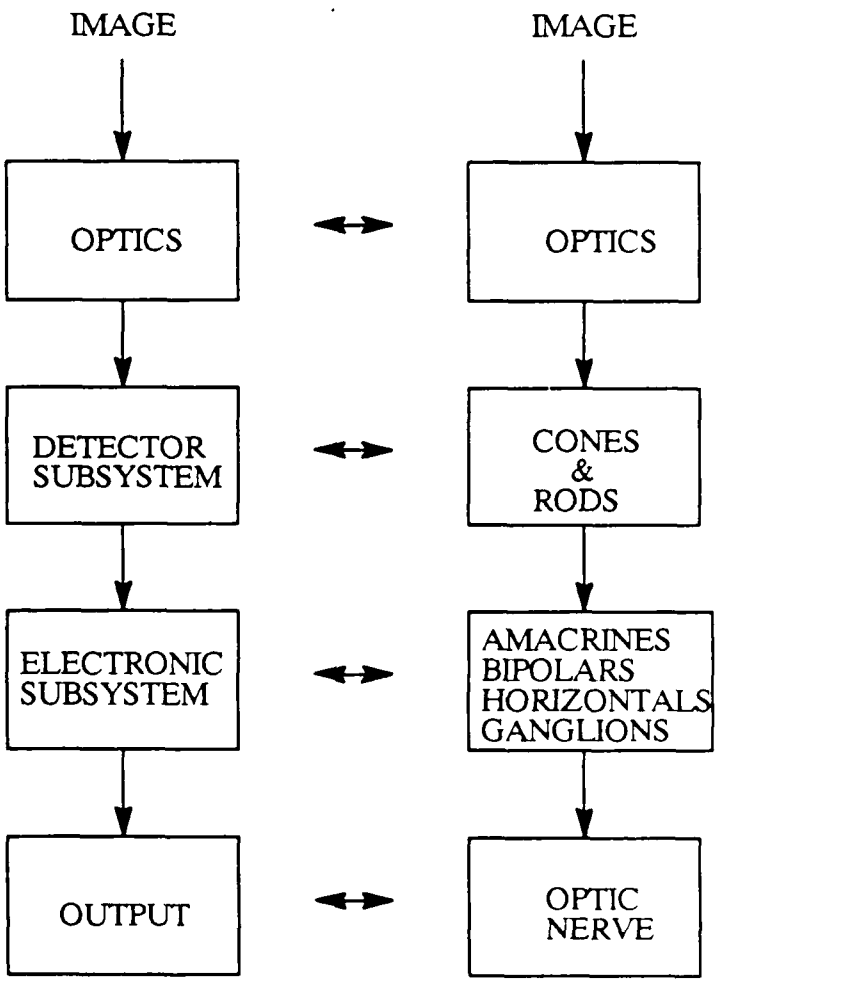


Figure 2-5: VISUAL SYSTEM DESIGN

2.3.1 Optical subsystem. The optical subsystem consists of those functional components (lens, mirrors and filters) which focus an image onto the detectors. The optical design is beyond the scope of this effort though assumptions have been made to use in the calculations on the focal plane design. The biological system, which for young adults can focus from 12 cm to infinity, is modeled by the following two parameters. First, the iris which serves as an aperture stop in the eye varies from a diameter of 2 mm in bright light to 8 mm in darkness, and second, the combined lens optical center of the eye is approximately 17.1 mm in front of the retina [Hec87].

The limit of resolution of an imaging system may be defined as:

$$(\Delta l)_{\min} = 1.22 f \frac{\lambda}{D} \tag{2-4}$$

where

(Δl) is the center-to-center separation of the images on the focal plane;

f is the focal length;

λ is the light wavelength;

D is the lens diameter [Hec87].

This equation provides a simple approximation of the minimum resolution of a system, which is the minimum separation of two images so they may be just resolved. The detector spacing on the FPA must be related to the Δl of the images to determine overall system resolution, because the detector spatial resolution provides a minimum resolution also. The Δl must be greater than twice the detector spacing to insure adequate sampling. Therefore, either the optical geometries or the detector spacing will determine resolution.

2.3.2 Detector Subsystem. The detector subsystem functional component is responsible for converting incident radiation from an image into current or voltage which may be processed by electrical circuits. The subsystem usually consists of an array of detector elements known as pixels, and when physically located in the focal plane of an imaging system is known as a focal plane array. The desired wavelengths to be detected determine many of the parameters of the detector. Detectors are commercially available for wavelengths from the ultraviolet to the far infrared [Wya87]. This discussion of detectors concentrates on the visible range which is the range the retina operates in, but the discussion aptly applies to any wavelength.

2.3.2.1 Detector Figures of Merit. A detector may be characterized by spectral response, quantum efficiency (η), responsivity (R), noise equivalent power (NEP), and Detectivity (D^*).

The spectral response of a photo sensor is largely material dependent. A material that responds well to ultraviolet wavelengths may be unresponsive to infrared and vice versa. Therefore, the system spectral response requirement will determine to a large extent what material the photo sensor is fabricated in. Appendix B contains a nomogram which shows the wavelength of different electromagnetic regions as well as the relation between wavelength and photon energy (in electron volts). Photon energy may be computed by the equation:

$$E_o = h\nu \tag{2-5}$$

where

h = *planks constant*;

ν = frequency.

Wavelength is related to frequency by the equation:

$$\lambda = \frac{c}{\nu} \tag{2-6}$$

where

c = speed of light;

λ = wavelength;

ν = frequency [Ros79].

The material also determines the absorption coefficient (α). The absorption coefficient is largely dependent on wavelength, and it determines the range of wavelengths for which photo current may be generated (those in which the photons become absorbed) [Sze81]. Appendix C gives the absorption coefficient for various photo detector materials.

The quantum efficiency is the number of electron-hole pairs generated per incident photon. An ideal detector would have a quantum efficiency of one (100%). The following equation describes its relationship to incident illumination:

$$\eta = \frac{h\nu I_p}{qP_{opt}} \tag{2-7}$$

where

I_p is the photo generated current;

P_{opt} is incident optical power at some wavelength λ;

q is elementary charge;

$h\nu$ is the photo energy corresponding to λ.

Responsivity, the ratio of photocurrent to the optical power, is related to quantum efficiency. The following define responsivity:

$$R=\frac{I_p}{P_{opt}}=\frac{\eta q}{h\nu}=\frac{\eta\lambda(\mu m)}{1.24} \tag{2-8}$$

The responsivity increases linearly with wavelength for a given quantum efficiency [Sze81].

The noise equivalent power (NEP) is the amount of incident light, that would generate a photo current equal to the dark current. The dark current is shot noise (the random fluctuation in photocurrent due to the random arrival of photons), so the NEP determines the lowest level of light which can be measured [Ros79]. The noise equivalent power of a 100% modulated signal is given by the following equation:

$$NEP=2(\frac{h\nu}{\eta})(\frac{I_{eq}}{q})^{1/2} \tag{2-9}$$

Where

I_{eq} = is the equivalent current defined by;

$$I_{eq}=I_b+I_p+\frac{2kT}{qR_{eq}};$$

and where

I_b = Background current;

I_p = Dark current;

R_{eq} = the equivalent resistance of the detector;

T = Absolute temperature;

k = boltzmann constant [Sze81].

Detectivity, the most used figure of merit for infrared detectors is given by:

$$D^*=\frac{A^{1/2}B^{1/2}}{NEP} \tag{2-10}$$

Where

B = the bandwidth;

A = Area of the Detector [Sze81].

In practice the D^* may reference detectors that sense radiation from either a black-body source or monochromatic source so that should be delineated clearly. Sze recommends explicitly showing the source by using $D^*(\lambda, \nu, 1)$ for monochromatic source and $D^*(T, \nu, 1)$ for a black body source [Sze81]. The higher the detectivity, the lower the radiation level to which a detector can respond. High detectivity is therefore desirable.

2.3.2.2 Photodiodes/Phototransistors. Light effects all semiconductor diodes and transistors to varying extents, which is the reason the packages which contain the electronic devices are designed to block light. Photodiodes and phototransistors sensors are specially designed to efficiently convert light energy to electric current, and the packaging illustrates this. These devices are individually addressable and thus impose resolution constraints for even small arrays, because the metal address lines decrease the percentage of photosensitive area.

2.3.2.2.1 Photodiode. The photodiode is a P-N junction operating under reverse bias conditions (Figure 2-6). Light striking the P-N junction generates electron-hole pairs in the depletion area. The reverse bias sweeps the electrons away, generating current. The size of the depletion region determines the transit time and the quantum efficiency. A thin depletion region reduces the transit time (because there is less distance for electrons to flow) while reducing the quantum efficiency (because a smaller fraction of the incident light is absorbed). A thick depletion region has the opposite effect. The thickness of the depletion region is determined by doping levels; the higher the implant the thinner the depletion region. Therefore, the system's purpose will determine this parameter [Sze81]. Additionally, the diodes are generally kept small to minimize junction capa-

citance.

2.3.2.2.2 Phototransistors. Phototransistors function like photodiodes but amplify the current produced. In the phototransistor, light striking the floating base-collector junction creates the electron-hole pairs. This turns on the transistor and an amplified current flows from the emitter to the collector. Phototransistors typically have higher capacitance than photodiodes so their high frequency performance is lower.

The CALTECH design takes advantage of the inherent parasitic bipolar transistor in CMOS [Mea87]. Normally a source of latchup, this device also may serve as a good phototransistor.

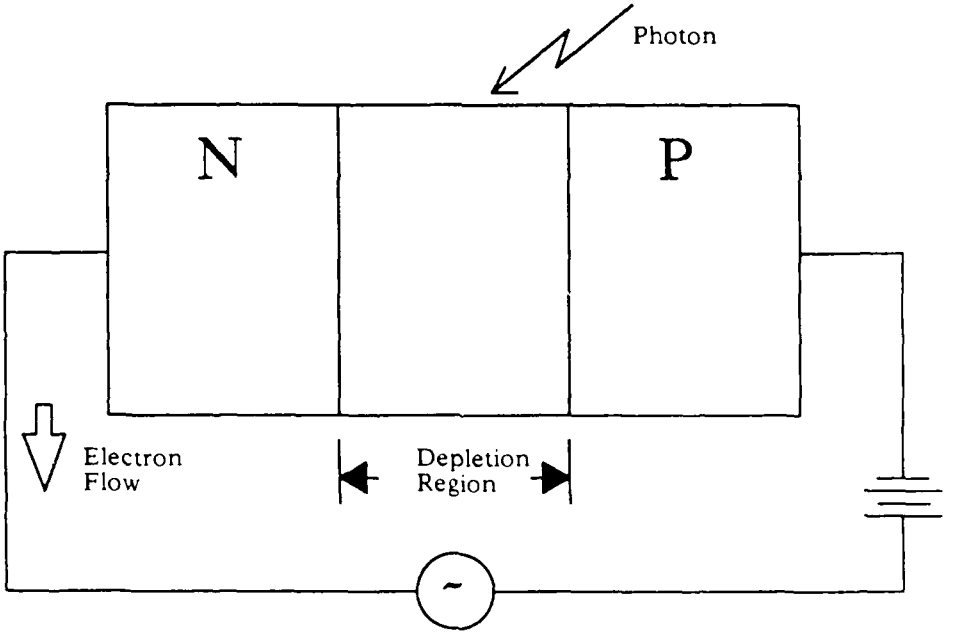


Figure 2-6: PHOTODIODE SCHEMATIC

2.3.2.3 Charge-Coupled Devices (CCD's). CCD's provide an advantage over photodiodes/phototransistors in large arrays by solving the readout problems. CCD's do not require individual select lines or read out wires so there is a higher percentage of photosensitive area. The CCD operates on the "bucket brigade" principle. Charge packets are transferred from one photogate to another, under control of a series of clock pulses. The basic concept is show in Figure 2-7. A prime concern is the transfer efficiency of the CCD. That is, how much of the charge packet is lost moving from one well to another. In practice, transfer efficiencies higher than 99.9% are routinely obtained [Sze81]. CCD arrays inherently require sequential access.

2.3.2.4 Charge-Injection Devices (CID's). The CID is a hybrid between CCD's and photodiodes. Each photo cell may be individually accessed, however, metal lines are not necessary to accomplish this. Therefore, the pixels may be packed in dense arrays providing a high percentage of photosensitive area. Depending on the device design, readout of the photocell is accomplished by either measuring the charge that flows upon injection to the substrate or by measuring the change in voltage induced when the charge is transferred from one cell electrode to the other. Figure 2-8 shows the operation of a row readout CID, an example of the latter method of CID readout. In this device, the horizontal scanning register sequentially connects the array columns to the column drive voltage. This forces the charge in each cell of the selected column to transfer to the row electrode where the change in voltage is sampled when selected by the vertical scanning register. At the end of the line scan, both column and row electrodes can be driven to inject the charge into the substrate to clear the cell.

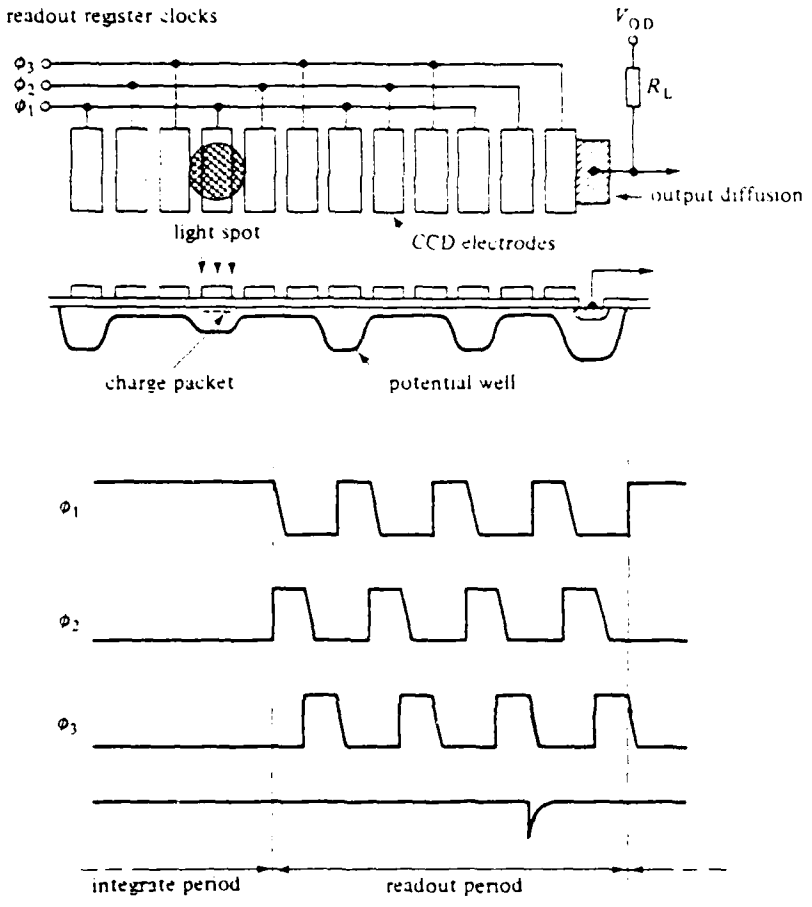


Figure 2-7: CHARGE-COUPLED DEVICE
(from [Bey80])

2.3.3 Electronic Subsystem. The electronic subsystem includes electronic components which perform signal conditioning and low-level processing on the detector outputs. Specifically, the electronic subsystem performs the retinal functions of mutual inhibition, spatial processing, and temporal processing. The CALTECH 2D electronic retina design is used for modeling purposes.

2.3.3.1 Logarithmic Amplification. The retina uses a logarithmic scale to report incident intensity [Wer73]. This helps compress the wide range of input intensities into a smaller output range for processing. Much the same way, the CALTECH retina uses a logarithmic amplifier to compress the output current of each pixel to a voltage range within CMOS voltage range.

2.3.3.2 Mutual Inhibition. In the CALTECH design mutual inhibition. described earlier, is implemented by using a resistive array. The sensors are arrayed in a hexagonal pattern with resistors connecting the nearest neighbors. Each pixel provides

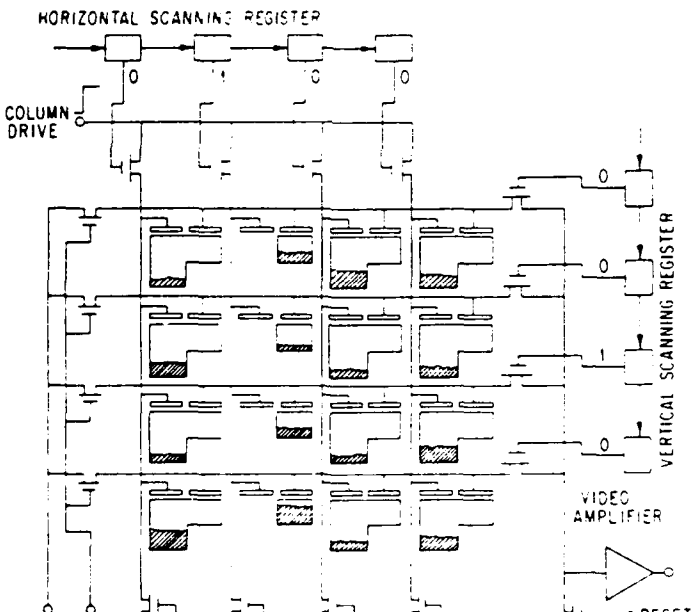


Figure 2-8: OPERATION OF A CID
(from [Mic80])

input to the resistive network, thus having an impact on all other pixels, which decreases with distance because the value must traverse more resistors. The net effect is a spatial average of intensity with more influence exhibited by nearby pixels. The result is a very effective local gain control [Mea87].

The resistive network is fabricated using an array of transistors. The MOS transistors are operated in the linear region to act as resistors (see Figure 2-9). There is a certain amount of fault tolerance built into this arrangement. If a pixel becomes stuck, the current is limited so the entire system does not crash. The resistive array is analogous to the horizontal cells in the biological retina [Mea87].

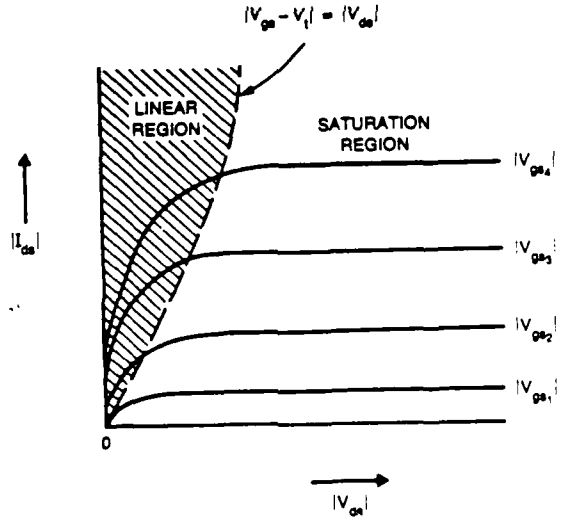


Figure 2-9: LINEAR REGION OF MOS TRANSISTOR
(from [Wes85])

2.3.3.3 Spatial Processing. Spatial processing takes place in conjunction with mutual inhibition. The net result of spatial processing is an enhancement of contrasts. In the CALTECH design, MOS transistors are used to form a differential amplifier. The difference between the spatial average which was computed as described above, and the local signal is computed. This spatial derivative is used as the signal corresponding to the output of the bipolar cell in the retina [Mea87].

2.3.3.4 Temporal Processing. Taking the derivative with respect to time of the output of the spatial processor provides temporal processing. Transistors are used to implement two amplifiers and a capacitor [Mea87]. The resulting circuit performs somewhat similar to the Amacrine cell (in that motion is detected). The net effect of this process is an edge enhancement of a moving object (see Figure 2-10). The output from this cell is amplified and sent to a monitor for viewing, at some determined rate. In the biological retina this signal would be input to a ganglion cell, where the signal would be converted to a constant amplitude, frequency varying signal.

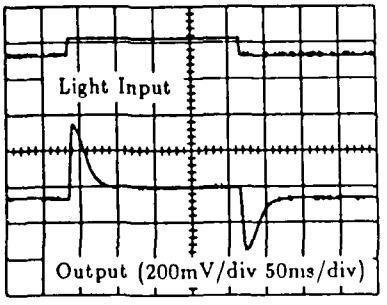


Figure 2-10: CONTINUOUS TIME DERIVATIVE
(from [Siv87])

2.4 Summary

This chapter presented the anatomy of the biological retina and its known functions. A systematic design procedure for electro-optics was presented and related to the biological eye. Several alternatives of photosensors were presented. The actual choice depends on the architecture and spectrum sensitivity needed. The electronic subsystem is the big variable. For this research effort, the subsystem is a replica of the CALTECH design. Even though it has certain limitations, it does significantly show the possible improvement in using analog processing elements as a data reduction technique over conventional digital techniques.

CHAPTER 3

Architectural Design

3.1 Overview

The material in the previous chapter showed the methodology and theory needed to emulate a retina with an electro-optic design. There are several architectures which could be used in the electro-optic design. These include varying combinations of focal planes and electronics in 2D and 3D which may be used in an electronic retina architecture. This chapter describes the system requirements and then outlines four general architectures which may be used to satisfy the requirements. Each has various advantages and disadvantages which are described. Finally, a comparison is made to select the architecture which best fits the system requirements.

3.2 System Requirements

The requirements for the electro-optic implementation of the retina are outlined in Table 3-1. These requirements were gleaned from the information presented in Chapter 2. The objectives are the general characteristics necessary to emulate the retina. The most important characteristics are real time processing and operation in the visible spectrum.

The operating environment gives those conditions which the design must be capable of handling and is the set of conditions which effect the biological retina. These include temperature, light sources, and relative movement between the scene and the focal plane.

Table 3-1: SYSTEM REQUIREMENTS

Objectives
- High Resolution Imaging
- Real Time Processing
- Visible Spectrum (390 nm to 780 nm)
- Long lifetime, small size, low maintenance

Operating Environment
- Ambient Environment
  - -- Earth atmosphere, uncontrolled, variable range
  - -- Ambient sources (sun, moon, lights)
  - -- Optical path: atmosphere
- Object Characteristics
  - -- Temperature: ambient
  - -- Shape/size difference
  - -- Spectral response: visible
- Background Characteristics
  - -- Terrain, atmosphere
- Overall Geometries
  - -- Range: Variable
  - -- Resolution: high
  - -- Total field of view: variable
- Sensor Scene Dynamics
  - -- Update rate: 100 Hz
  - -- Relative velocity: varying

Certain functions are done by higher level processors. For example, if the retina is moving in relation to the scene, then a correction will have to be computed. High-level processors will use the retina's relative velocity to accomplish this. Other constraints are highly dependent on the optics. For example, the field of view, range, and resolution are largely optics driven. The spatial separation of pixel elements contribute to this also.

There are two techniques used to provide the high throughput necessary to process the large data arrays (512 x 512 or larger). The first approach uses massively parallel processing. The problem, as alluded to earlier, is fitting the parallel computer into a small space. The 3D computer provides the means to fit a highly parallel, high throughput computer into a very small space. A 512 x 512 data path 3D computer will fit in a 1 liter container (about the size of a coffee can). This is small enough to be con-

sidered as useful in autonomous robots or in aero-space systems where weight and volume are a major concern.

The other technique is to eliminate computations. This is done through the use of analog circuitry replacing digital logic. The low-level processing algorithms used to detect edges, motion, and contrasts are computation intensive and must be done on all the pixels in the frame. Current algorithms require 100 to 500 mathematical operations per pixel per image frame to perform with digital processors. These are precisely the type operations continuously performed in the retina.

Assuming 200 operations per pixel per frame, 512 x 512 frame size, and 100 Hz frame rate; a serial digital visual system requires 5.24 x 109 operations per second (about 5.2 BOPS) for real time processing. For images requiring 500 operations per pixel per frame, 1.3 x 1010 operations per second (about 13 BOPS) are necessary for real time processing. And this is only the low-level processing. Heat sensors which operate with slower integration times are less constrained but still require significant throughput, while high performance visual systems may operate at 1000 Hz to 10000 Hz, requiring 1 to 2 orders of magnitude higher throughput than human rates. In any case, the retina using analog circuitry accomplishes this task. The CALTECH retina also accomplishes low-level processing in real time using analog circuitry. Through analog circuitry the intensive digital computations are eliminated; the low-level processed image is available at the frame rate (assuming separate analog circuitry per pixel). The analog output may then be converted to digital for the high-level processing necessary to analyze the image.

The high-level algorithms such as object identification may also benefit from highly parallel processing, though not to the same extent as low-level processing. The reason for this is the data reduction which results from the low-level processing. The information

in the scene is contained in the edges and in motion, so the high-level processor only operates on reduced data which it receives from the low-level processor.

System input/output constraints also need to be alleviated. As image arrays get larger, serial data I/O becomes a major bottleneck. A 512 x 512 array at a 100 Hz frame rate provides 26.2 M samples per second. Assuming an eight bit quantization (256 levels), this results in about a 210 MHz serial data rate. Under the same assumptions a 1024 x 1024 array requires 838 MHz. Alternatively, a 512 x 512 array with 512 data lines operates at about 410 KHz to transfer the data. Clearly, a parallel approach to some degree is preferable to accomplish real time processing as the array size increases. In most systems, the number of pins and wires for I/O is limited and tradeoffs need to be investigated. Two methods to circumnavigate this constraint include using the feedthroughs supported by the Hughes technology, or using optical interconnects. The Hughes technique provides additional bandwidth through parallel channels while the optical interconnect provides increased bandwidth on a single high speed (1 GHz and higher) channel.

3.3 Alternative Architectures

Any proposed architecture must perform analog low-level processing, and provide high throughput considering both I/O constraints and processing constraints. Obviously, there are many permutations of architectures which may be devised to accomplish this. However, four general architectures are presented which provide the basic advantages of most permutations. Design 1 physically separates the focal plane and low-level processor from the high-level processor. Design 2 uses the epitaxially grown 3D circuit technology. Design 3 places the focal plane/processor as the top wafer in the Hughes 3D computer architecture. Finally, design 4 is a novel approach which uses fiber optic interconnects

between stacked wafers.

For purposes of comparison, the sensor/low-level processor element from the CALTECH design is used as the primitive cell in the array. This could be replaced by any combination of pixel and analog processing cell, being careful to match area requirements. The primitive cell has a size of 164 x 145 microns using 3 micron feature size, or about 72 X 70 microns in 1 micron technology with the same size sensor element. The photosensitive area of each pixel is $675\mu m^2$ for both technologies. The 3 micron technology, with about 3% of each cell photoactive, has very poor resolution and is inadequate for most visual processing. The 1 micron technology with 13% photoactive area allows for adequate low resolution processing.

3.3.1 Architecture Design 1, Separate FPA. This architecture follows conventional designs with the FPA on a separate chip. The FPA consists of a 512 x 512 array of primitive cells. It breaks with conventional designs as low-level processing is done on chip with its collocated pixel in the primitive cell. Figure 3-1 depicts this architecture. Because I/O is accomplished conventionally (pins), no feedthroughs are required in this design. The 512 x 512 array of elements (1 micron technology) uses a square area approximately 1.5 inches on a side.

A truly conventional design would use an array of CCD's and a serial output. However, the advantages of CCD's are not useful in this case due to the non-contiguous nature of the pixels. As described earlier, with the use of 1 micron technology, 13% of the FPA is photosensitive. A pixel spacing of 40 microns provides a spatial resolution adequate for low resolution visual purposes. A conventional CCD array would provide higher resolution capability, however, this would necessitate separating the photosensitive areas from the processing elements. The image would have to be sampled prior to

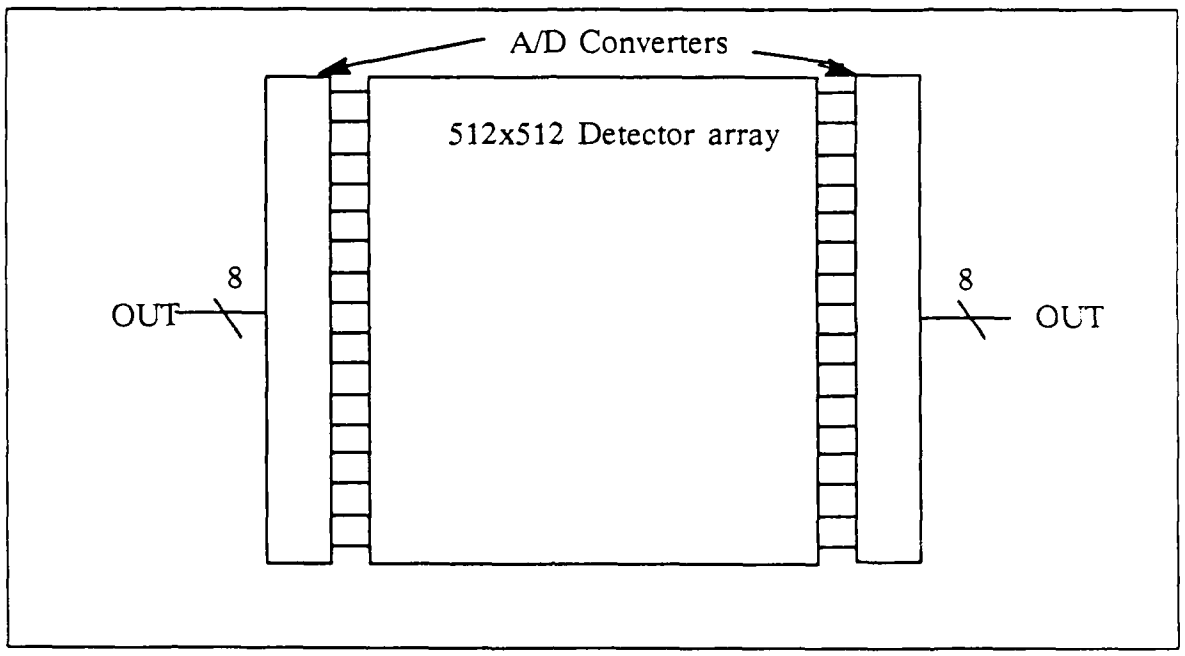


Figure 3-1: ARCHITECTURE 1

low-level processing and shifted to the processing element array, adding a frame delay in the whole process because the analog cells require the whole image value to work accurately. Furthermore, the problem is then transformed to the more computationally complex correspondence problem (determining what point in the present image frame corresponds to what point in the previous image frame).

A major advantage of this design lies in the ability to easily replace the FPA. Since the focal plane with its sensors is generally in a more exposed position than the processor, it is more likely to be damaged. Therefore, being able to replace a chip or wafer FPA is more advantageous than replacing a whole FPA and processor assembly. This

architectures allows schemes to be devised with backup FPA's feeding the same processor. Also, since the FPA is separate from the larger high-level processor it will fit in a smaller area.

The major disadvantage lies in the necessity to transmit data off the wafer to a high-level processor. For small array sizes this does not present a problem. However, as array sizes become larger, 512 x 512 and up, the I/O becomes a major constraint.

3.3.1.1 Operation. The image is focused on the FPA where the image is sensed and low-level processing is accomplished. The results are converted to digital using two analog to digital converters. With a 100 Hz frame rate and a 512 x 512 array, this results in 13.1 M samples per second per converter. The array is split, with half the columns (256) going to each A/D converter. Each converter operating at approximately 13 MHz provides the required 8 bit quantization. The output of the chip consists of two pixel streams of 8 parallel bits each. The output image is mapped into a high-level processor at the 100 Hz rate.

3.3.2 Architecture Design 2, Monolithic 3D Fabrication. This architecture utilizes the 3D circuit fabrication process in which a second layer of circuits is fabricated on top of a completed layer. The top layer consists of the photo sensor array and the "buried" circuits consist of processing elements. Figure 3-2 depicts this architecture. Note that the top wafer of the 3D stack now consists of what normally would have been two wafers. Ideally, there would be more than 2 levels of circuits and the high-level processing could also take place on one multilevel monolithic wafer, negating the need for stacked wafers. However, in this design with only 2 levels in the wafer, the processing elements again are laid out in a 512 x 512 array of cells, each with its own vertical

feedthrough to the next lower wafer, where higher level processing begins.

Because there is no feedthrough separation, the processor elements may be more densely packed providing good vertical registration from each pixel element to its processor. Using the previous cell design as an example, the photodetectors need a 33 x 33 micron square (of which only $675\mu m^2$ is photosensitive). The processing elements directly below will fit in a 60 x 60 micron square. Since only the sensors are on the top of the wafer, 19% of the FPA area is photoactive with a 27 micron pixel separation. With this design though, the sensor size can be increased to almost the size of the

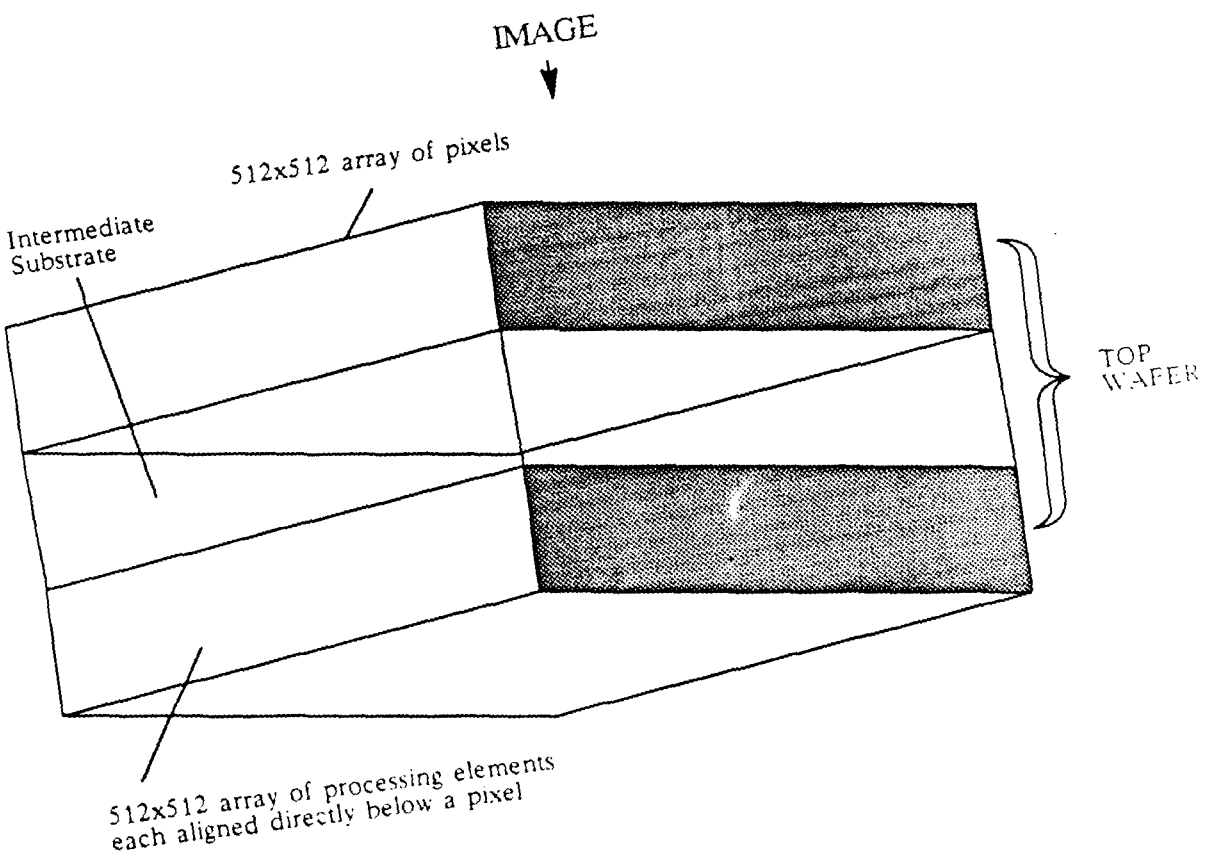


Figure 3-2: ARCHITECTURE 2

underlying processor cell. The result will be a near contiguous array of sensors with 90+% photosensitive area on the FPA. This would provide extremely high resolution similar to that achievable by CCD arrays.

Another consideration is the approximately 4 mil square feedthrough area in each of the 512 x 512 cells of the 3D computer. Wherever a feedthrough exists, there cannot be any electronic devices. Thus photo sensors cannot be located over feedthroughs. This results in missing pixels, or holes in the image. If there were more monolithic circuit levels available and/or the Hughes 3D structure was not used for high level processing, this problem would not exist.

A technique which solves this problem is to start with a 3D wafer without feedthroughs under the photo sensor array. The photo sensors and the processors are fabricated without regard to the cell array structure of the 3D computer. The output of each processor must be multiplexed to the periphery of the wafer, where feedthroughs were placed, similar to design 1.

3.3.2.1 Operation. This architecture is conducive to analog processing due to the short direct connects between pixel element and its processor. The image is focused onto the focal plane where current is induced. The current is then directed into the processing element directly below it by the vertical metal wires. The output of each processing element is then multiplexed to the periphery where feedthroughs provide a data path to the high-level processing wafers below (3D computer). Ideally the high-level processor would have been epitaxially grown also, providing a base for the FPA fabrication so no multiplexing or feedthroughs would be required. From this stand point, the whole system from pixels to high-level output could be viewed as one wafer. However, due to yield considerations, heat dissipation and other problems, this is not technically feasible for

large arrays at this time.

Problems with this approach stem mainly from the 3D circuit fabrication. The circuit must be fabricated in the same material as the substrate. This limits the flexibility of the spectrum sensitivity of the photo sensors to the substrate material. More importantly, this is a very immature technology and would result in low yield. Especially when coupled with the 1 μm VLSI on a wafer-scale level.

3.3.3 Architecture Design 3, Hughes 3D Computer. This architecture, as depicted in Figure 3-3, has the processing elements and photo sensor array on the same wafer in the focal plane. The wafer constitutes the top wafer of the 3D computer wafer stack. The feedthroughs are absent in the central portion of the chip where a 512 x 512 array of pixel/processor cells are located. Using 1 micron technology the array uses approximately 2.25 square inches (1.5 inches x 1.5 inches) in the center of the wafer. The 512 x 512 3D computer uses 6 inch wafers, so there is ample area on the wafer for A D converters. The 3D computer structure is arranged in the stacks underneath this wafer. The most suitable pixel element in this arrangement is a photo diode/transistor.

A major advantage of this design is that the technology is availability, at least in smaller array sizes. The 512 x 512 array size still requires some development work. A disadvantage is that the technology to fabricate the 3D computer is not public domain

3.3.3.1 Operation. The operation of this chip is similar to architecture design 1. However, the vertical bus is constrained to 10 MHz. Therefore, the quantized samples cannot be converted at a 13 MHz rate as a bottleneck will develop. To perform at the 10 MHz rate, the array is partitioned in quarters. Each quarter has a dedicated 10 MHz A/D converter. The output of each is placed on 8 vertical lines and mapped to a 512 x

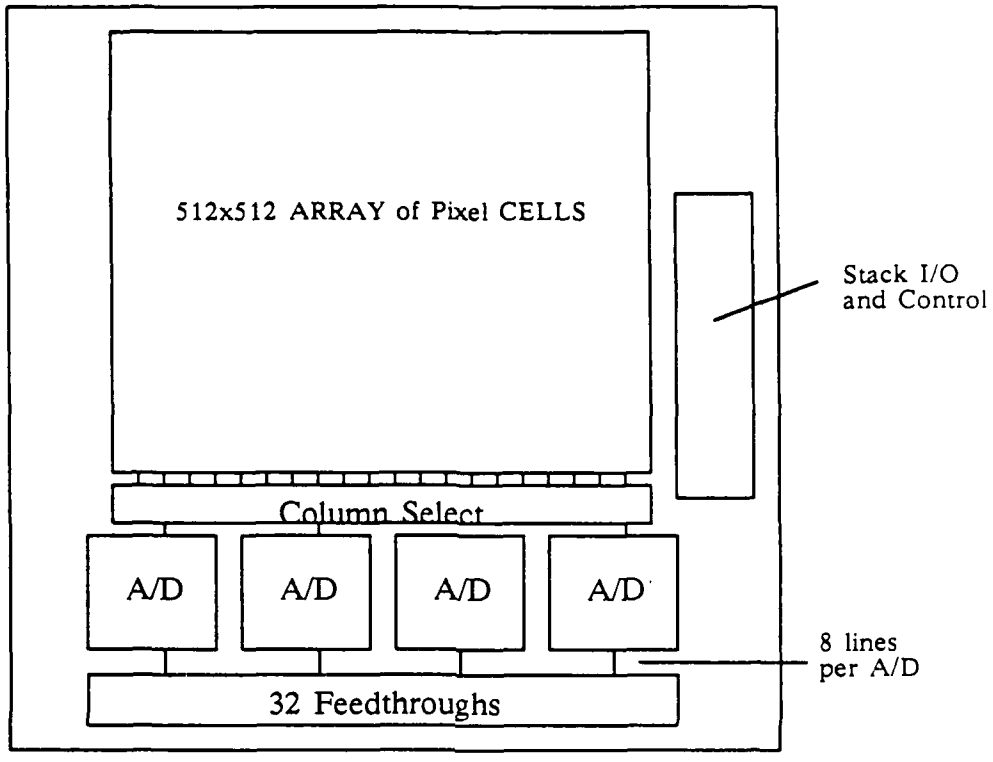


Figure 3-3: ARCHITECTURE 3

512 array of buffers on the next wafer down. This wafer is accessed by the high level 3D computer consisting of the remaining wafers in the stack. The 3D computer then operates on the data in SIMD fashion with a new image frame being supplied at 100 Hz. With the general architecture as outlined, for slightly faster performance, a 152 Hz frame rate could be supplied with no changes.

3.3.4 Architecture Design 4, Optical Interconnect of Stacked Wafers. Figure 3-4 depicts this design. It consists of a wafer containing the 512 x 512 array of primitive cells in the middle of the wafer. A GaAs chip containing a 210 MHz (to provide bandwidth for 100 Hz frame rate with 8 bits per pixel in the array) optical interconnect interface is attached to this wafer. There are 4 10 MHz A/D converters which feed the optical interconnect. The optical fiber terminates at the next lower wafer in the stack, where high-level processing is initiated. This next wafer could be a wafer in the 3D computer with feedthroughs or a series of wafers with optical interconnects. The fiber cable is routed through a hole drilled or etched in the substrate, or is routed along the wafer edge.

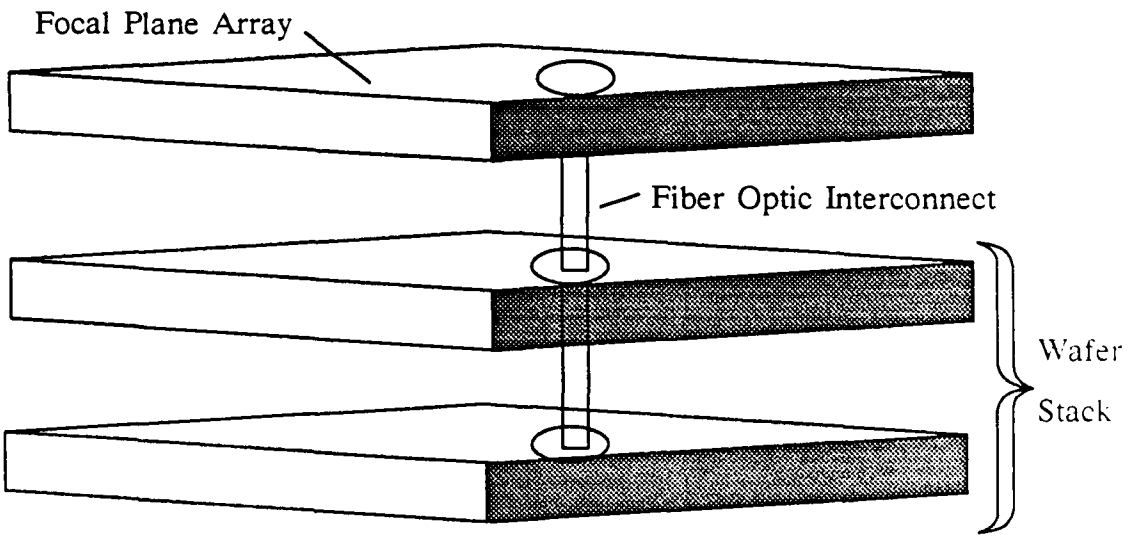


Figure 3-4: ARCHITECTURE 4

3.3.4.1 Operation. Similar to the other designs, the FPA is illuminated with an image. The low-level processed image is read out at a 100 Hz frame rate. To accomplish this, each column is partitioned into fourths. As a column is selected, each partition unloads its cell charge in a CCD shift register. The columns are selected at a 51.2 KHz rate. The charges are linearly shifted to its corresponding A/D converter at 6.6 MHz. The outputs of the A/D converters consist of 8 bit quantization on parallel lines. The four outputs are multiplexed through the optical interconnect at the required 210 MHz rate.

Disadvantages with this design stem from the high complexity of the hybrid design. Different supply voltages are required as well as different interface characteristics. An advantage is that the technology is not proprietary and thus not constrained by patents. Also, if higher frame rates are necessary, optical interconnects can be designed in the GHz range or multiple optical interconnects can be considered.

3.4 Comparison

The four designs outlined must be judged by how well they meet the system requirements. Foremost, is the necessity to operate in the visible spectrum. Through the use of silicon detector elements all the designs adequately accomplish this. Another important point is analog processing. The eye operates in analog, and there are many advantages for the architecture to support this type processing, besides just direct emulation. Through analog processing, real time processing, another major requirement is met.

The computing power necessary for low-level processing algorithms has been shown. The simple analog processing circuits perform the same computations in a fraction of the time. The output of the analog processors result in a data reduction as only the essential

(edge/motion) image information is sent to the high level processor. Therefore, architectures which best support analog processing are desirable. Furthermore, it is desirable to take the time derivative of the scene prior to sampling so as not to lose the most important data (the directional derivative with respect to time) in the image [Mea87]. All the architectures support this.

A major consideration in the selection of an architecture is the input/output constraints. The retina has an extremely parallel I/O architecture. For real time processing this is a necessity. The CALTECH retina used serial I/O to minimize output pads and devices. At 100 Hz rates and small array sizes, this may be adequate. However, with large arrays I/O times are significant. The optimum design would support a sole connection from pixel to processor to output. Only architecture 2 supports this and it is not currently feasible for large arrays.

The next best design would allow for a large percentage of the data to be transmitted at a time. All the other designs provide some means to provide data at visual processing rates. Each has advantages and disadvantages. Unlike design 2, all are based on a common primitive cell. Therefore, the differences lie mainly in the methods of I/O connections. The I/O constraint is manageable for all the designs with 512 x 512 arrays. However, the retina has 125 M receptors and a 1000 x 1000 array of outputs to the brain. Eventually, to fully implement a retina the architecture must be capable of dealing with scaled up data rates corresponding to the larger array size and still fit in a small package. The 3D computer offers a compact unit to provide the high-level processing. However, using optical interconnects between stacked wafers could also provide a similar compact computer, though more complex. Designs using conventional interconnects will not be feasible for these large arrays, so design 1 is eliminated.

Another important system requirement is high resolution imaging. Using 1 micron VLSI technology provides only low resolution capability because of the low percentage of photoactive area on the FPA. The only ways to increase the percentage of photoactive area are to increase the sensor size (thus increasing the whole cell size), or eliminate all or some of the processing elements on the FPA allowing a higher density.

Increasing the cell size is a self defeating proposition as a major requirement is to miniaturize the system. Eliminating the processing elements is also unacceptable (unless they can be located vertically) as they are already at the functional minimum required to do contrasting and motion detection. Design 2 provides the added flexibility to separate the processing elements from the the FPA, yet remain collocated with the photodetectors (by being positioned directly underneath). Thus, the sensor size may be increased providing a high resolution mosaic even if the underlying circuits use 3 micron technology.

3.5 Summary

The system requirements of the electronic retina necessary to emulate a biological counterpart mainly center around real time processing of visible light in a compact package. Four architecture to implement the retina were described and compared. The fundamental strategy was to provide real time processing capability on two dimensional images. Maximum parallel I/O, very high speed serial I/O, or some combination of the two and maximum parallel processing is required to achieve this. The use of the 3D computer provides the maximum parallel processing capability. The 3D epitaxial circuit fabrication provides the closest approximation to the retina. Unfortunately, it will not be feasible for large array sizes for some time. The 3D computer with a top FPA is the next best approximation. However, using optical interconnects or a separate FPA provides the same functionality.

This leads to a paradox as to which architecture is the best to use now. Is it better to have a large array with low resolution in a compact package (design 3 and 4) or to have a small array with high resolution in a compact package (monolithic design)? The answer is the latter. Also, it may well be the case that large arrays in 3 micron technology are feasible in the monolithic structure before large 1 micron arrays can be reliably implemented in the other architectures.

CHAPTER 4

VLSI Implementation

4.1 Overview

The previous chapter outlined four architectures and evaluated their ability to meet system requirements. The results of the comparison dictates a two pronged approach providing an evolution path to the best retina architecture presented - the monolithic 3D design. The initial architecture uses the 3D computer architecture in which the top wafer (FPA) contains the photosensors and low-level processing elements. This design is presently achievable, at least in 128 x 128 array size. The final architecture will use the cells developed in the first design but will separate the detectors from the processors and implement them in a 3D monolithic wafer. This chapter describes the VLSI implementation of the initial architecture. The monolithic implementation follows from this with the exception of different detector cells. A separate detector cell is implemented for use in the final design.

The architecture follows the general architecture of the eye, with the parallel analog processing and a data reduction prior to high-level processing. Figure 4-1 shows the data flow in the architecture (both initial and final designs). The architectural difference between the two is the focal plane design where one has all the circuits on one plane; the other has two levels of circuits. Table 4-1 outlines the electronic retina characteristics of the initial architecture.

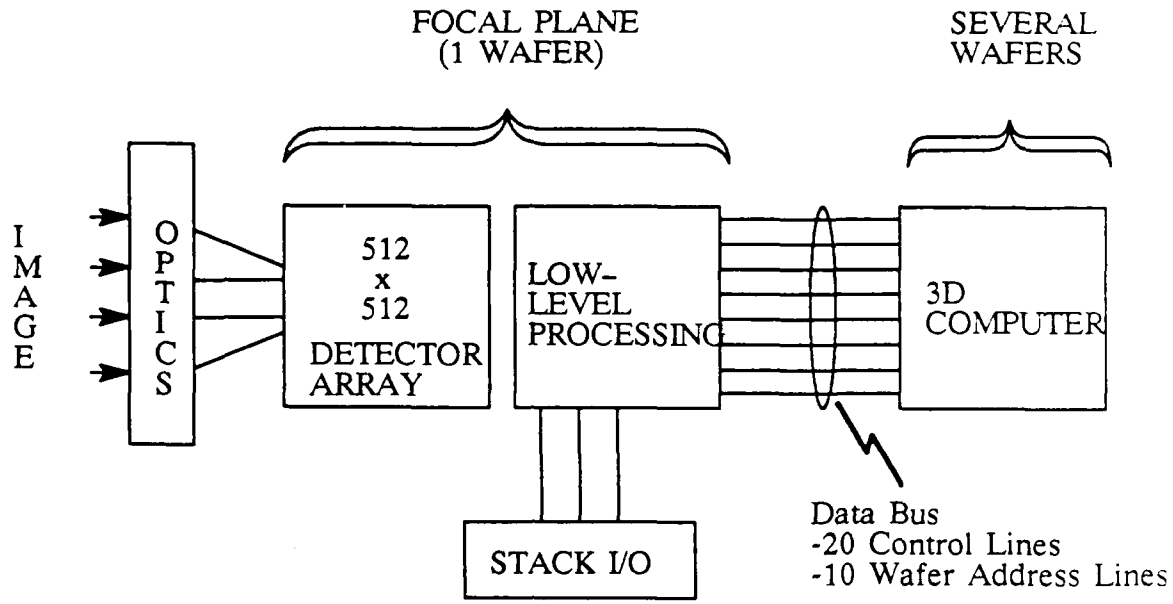


Figure 4-1: DATA FLOW IN THE 3D ARCHITECTURE

Table 4-1: 3D ELECTRONIC RETINA CHARACTERISTICS

Technology	CMOS
Packaging	3D Wafer Stack
Size	1 liter
Clock rate	10 MHz
Power Consumption	30 Watts

4.2 Design Rules

The design uses 3 micron p-well CMOS technology and lambda based (λ=1.5μm) design rules. The resulting design is scalable to smaller feature sizes to provide increased density and therefore higher resolution.

4.3 2D Implementation

The 2D implementation concerns the X-Y plane circuits on the wafer. The 2D cells implemented here are a CALTECH cell and a stand alone photodiode. The CALTECH cell is used in the initial design. The photodiode is an example of a detector which may be used on the top level of the monolithic architecture. The second level of circuits would be the CALTECH cell without the phototransistor. The other major cell type used is the 10 MHz analog to digital converter. This is a common implementation and nothing is special about it.

4.3.1 Retina Primitive Cell. The primitive cell (CALTECH cell), as previously described, consists of 4 major elements: phototransistor, logarithmic amplifier, time differentiator, and a spatial averager. A single cell CIF plot is shown in Figure 4-2. Figure 4-3 is a layout map corresponding to Figure 4-2 showing the relative position of the various elements in the design. The cells are laid out in a 512 x 512 hexagonal array to form the FPA. Figure 4-4 is a CIF plot of a sample array showing the hexagonal relationship of pixels.

The single cell has been fabricated to test for performance parameters. Five probe pads were positioned at various points in the cell to expedite thorough evaluation with a variety of input conditions. Figure 4-5 shows the cell with the probe pads inserted. The numbers on the lines correspond to pins on the package. Table 4-2 gives the pin definitions.

The differential amp number 2 bias (pin 7) controls the rate at which the the capacitor in Figure 4-6 charges. The bias on pin 9 controls the entire differential amplifier function. The readout select enables the sense amp and selects a row value to output on

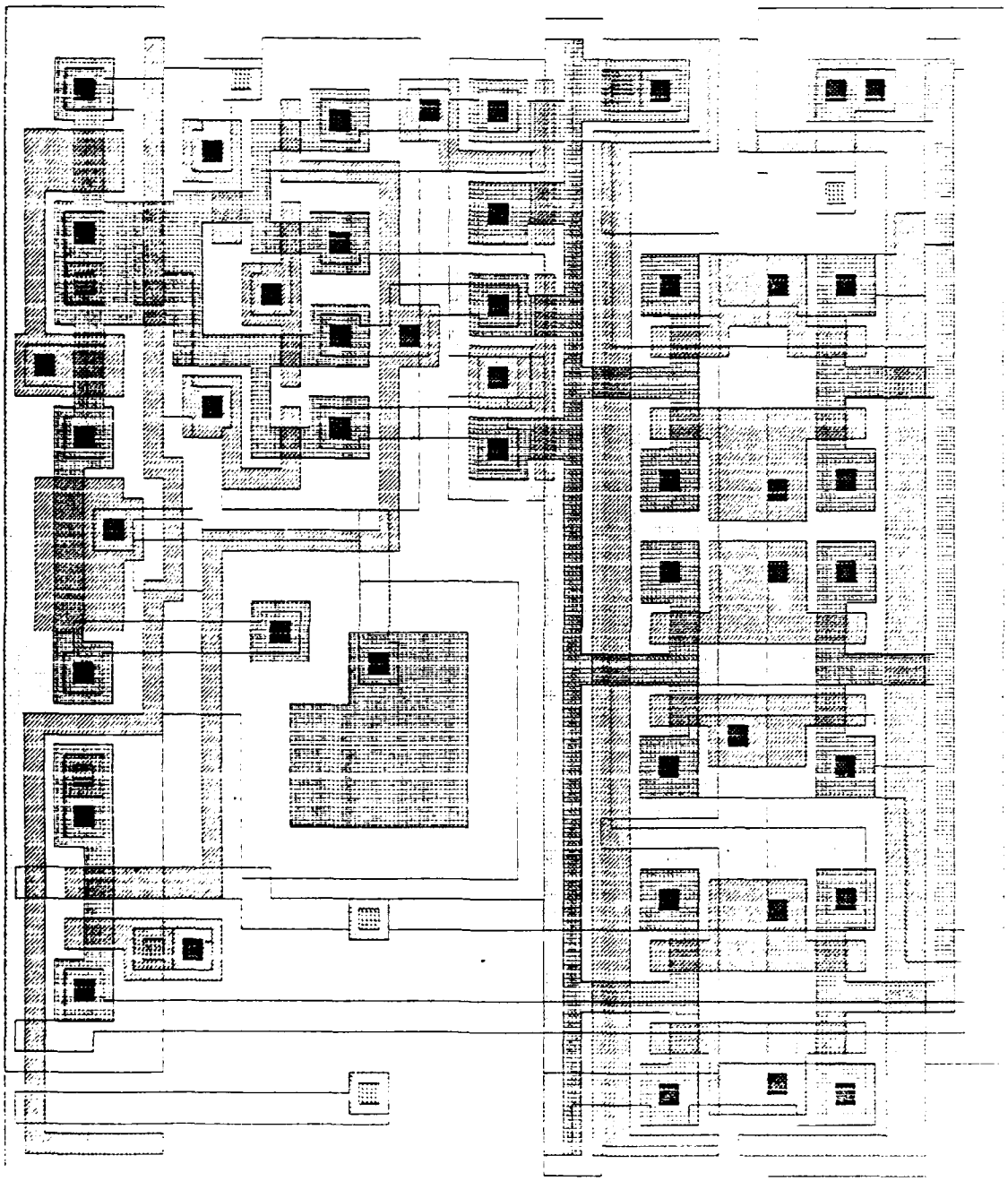


Figure 4-2: PRIMITIVE PIXEL/PROCESSOR CELL

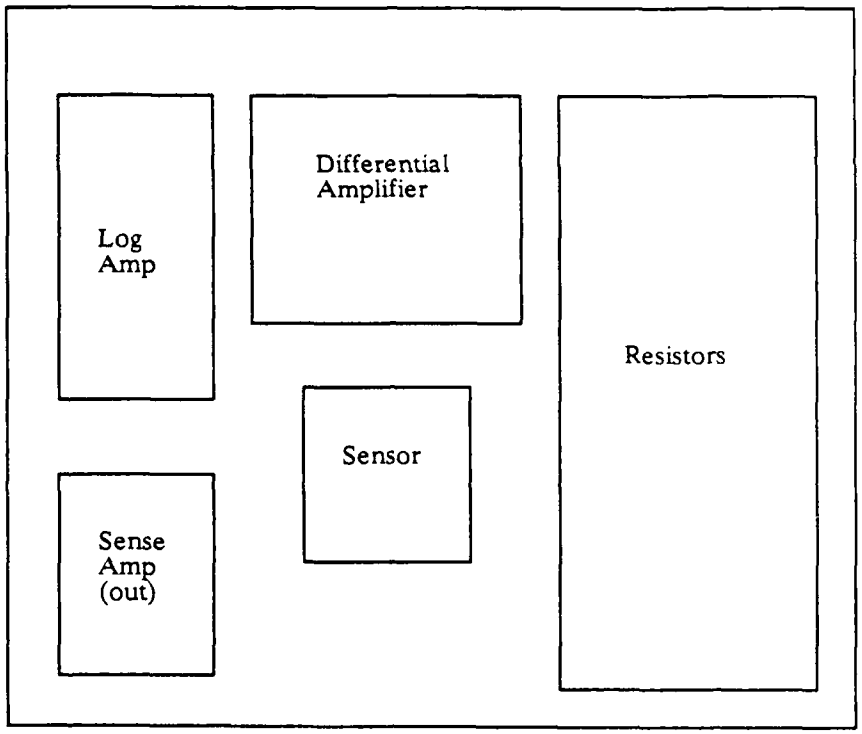


Figure 4-3: ELEMENT LOCATION

the column line. The direct output pin allows sampling of the cell output prior to the sense amp for testing purposes. Pins 14 and 15 provide the bias to operate the resistive network in the linear region. Pins 13 and 16-18 provide the means to directly read the resistor net value as well as put in various values to evaluate the spatial averaging function. Finally, pin 20 is the column readout where the cell value is directed to the analog to digital converter.

The differential circuit as shown in Figure 4-6 provides a continuous time derivative by comparing the present pixel value to some prior value. The prior value is a charge stored on the capacitor. If there is a difference between the two then there is either a positive or negative output change represented by a spike in the waveform graphed over time. This represents motion in the image at that particular point. Since the

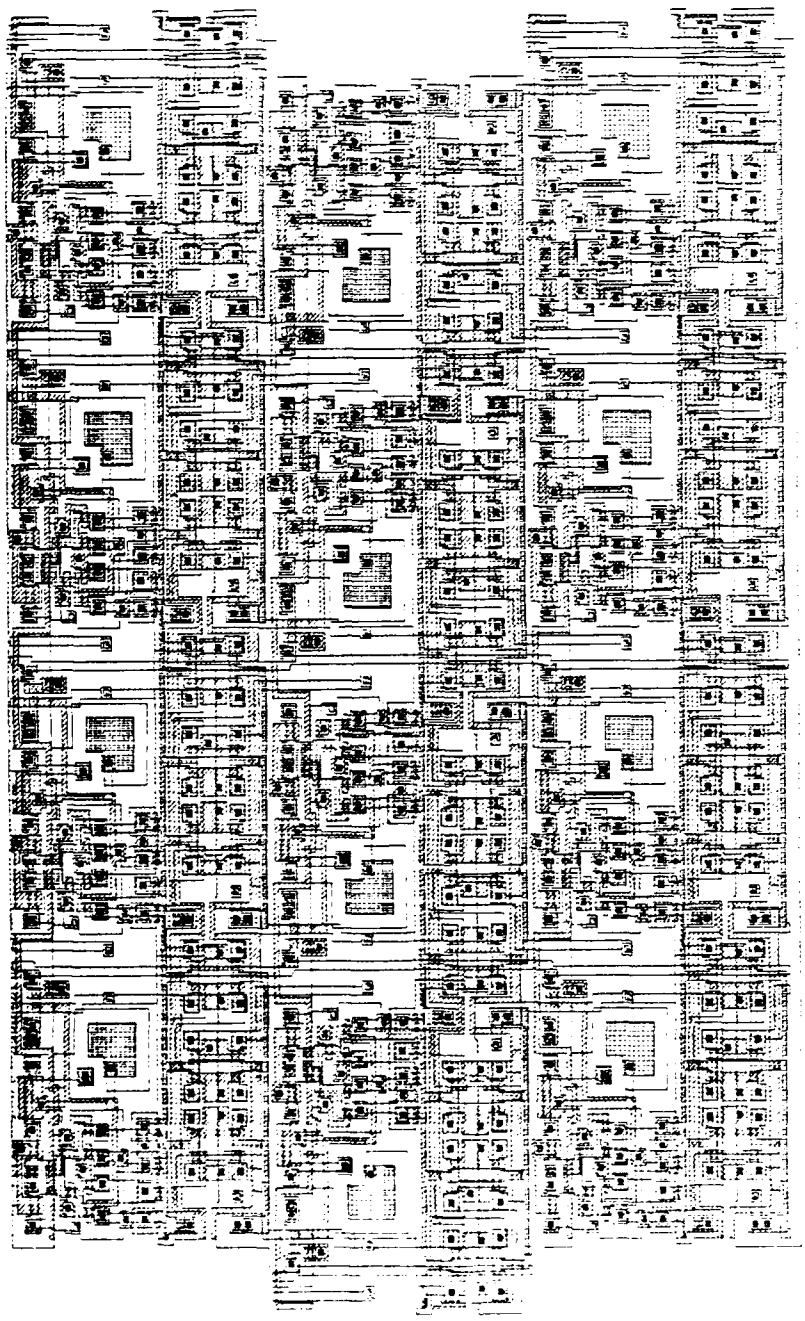


Figure 4-4: FPA HEXAGONAL ARRAY SAMPLE

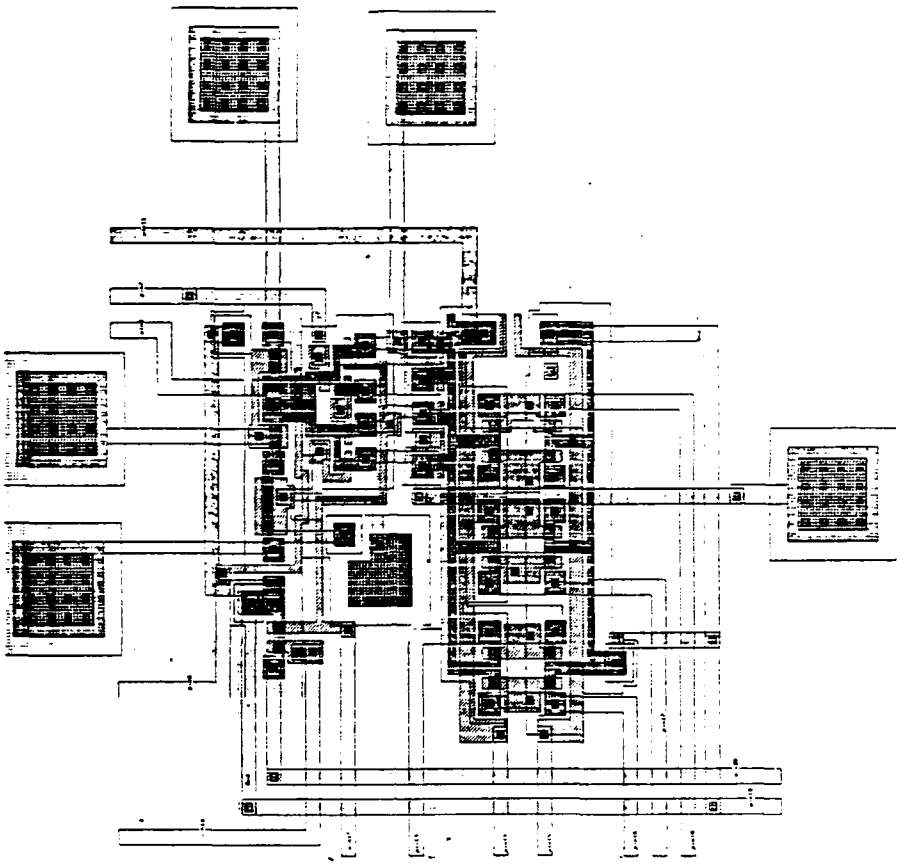


Figure 4-5: CELL SHOWING PROBE PADS

comparison is made with the spatial average (from the resistive network), motion is only detected in those areas which have high contrast (edges). Therefore, unless the resistive network is disabled, the output represents moving edges in the field of view. Also, if there is no motion, or the motion is such that a continuous edge is moving through the pixel, there is no output.

Figure 4-7 is a CIF plot of the bipolar phototransistor with the logarithmic amplifier separated from the rest of the cell. This cell was implemented to evaluate the luminance efficiency and the log voltage characteristics. There are 7 probe points to provide maximum testing capabilities. Only GND (pin 19) has been hardwired to this dev-

Table 4-2: PIN ASSIGNMENTS

Pin Number	Function
1	Substrate
7	V bias (Differential amp no. 2)
8	GND (log amp)
9	V bias (Differentiation control)
10	Readout select
11	Vdd
12	Direct Output
13	Resistor net total value
14	V bias resistors (p transistors)
15	V bias resistors (n transistors)
16	Local neighbor resistor input
17	Local neighbor resistor input
18	Local neighbor resistor input
19	GND
20	Output Column

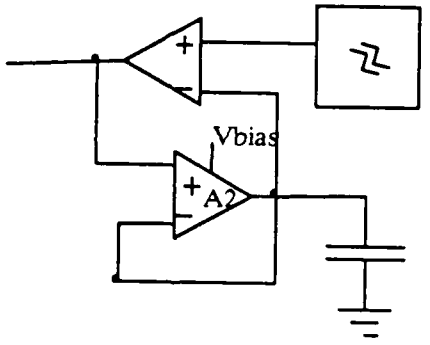


Figure 4-6: DIFFERENTIAL CELL DIAGRAM

ice. The purpose of the logarithmic amplifier is to compress the output current of the phototransistor into a voltage range compatible with the CMOS devices in the differentiator and resistor network.

4.3.2 Control Circuitry. The control circuitry consists of Vdd, Gnd, and 4 Voltage bias lines (see Table 4-2 and Figure 4-4). Three of the bias lines are in polysilicon and run the length of the array in 512 columns, so that every cell has access to the

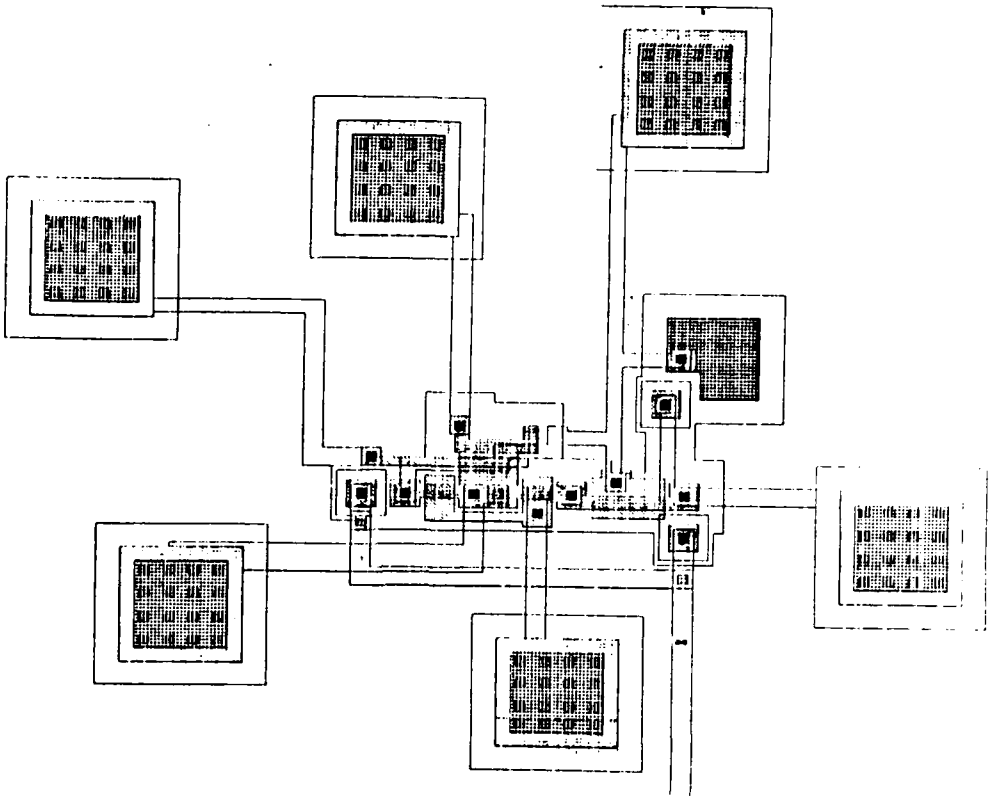


Figure 4-7: PHOTOTRANSISTOR AND LOG AMP

three lines. The other bias (differential amp number 2) is supplied in second metal which also has 512 columns the length of the array.

Vdd and Gnd are supplied to each cell by second metal which runs parallel to the long polysilicon lines. The second metal also provides a light shield to the non-photosensor active elements.

4.3.3 Input/Output. Each cell has a sense amp which is operated by a second metal select line. For a 100 Hz frame rate, this line is clocked at 51.2 KHz (512 row values times 100 frames per second). The output of the circuit is connected to a clocked multiplexor which feeds the cell output to an A/D converter. Since the columns are

logically partitioned in quarters, each quarter is routed to one of four A/D converters. Each column of the quarter is routed to the converter controlled by a 6.6 MHz (51.2 KHz x 128 columns) clocked shift register. The output of the converters are routed to feedthroughs. Figure 4-8 shows the I/O scheme.

4.3.4 Photodiode. Several photodiodes were fabricated to evaluate for use in the final monolithic design. The designs are similar to the vertical pn diode fabricated at the Jet Propulsion Laboratory, California Institute of Technology so as to be compatible with the MOSIS fabrication facility (see [Ber86]).

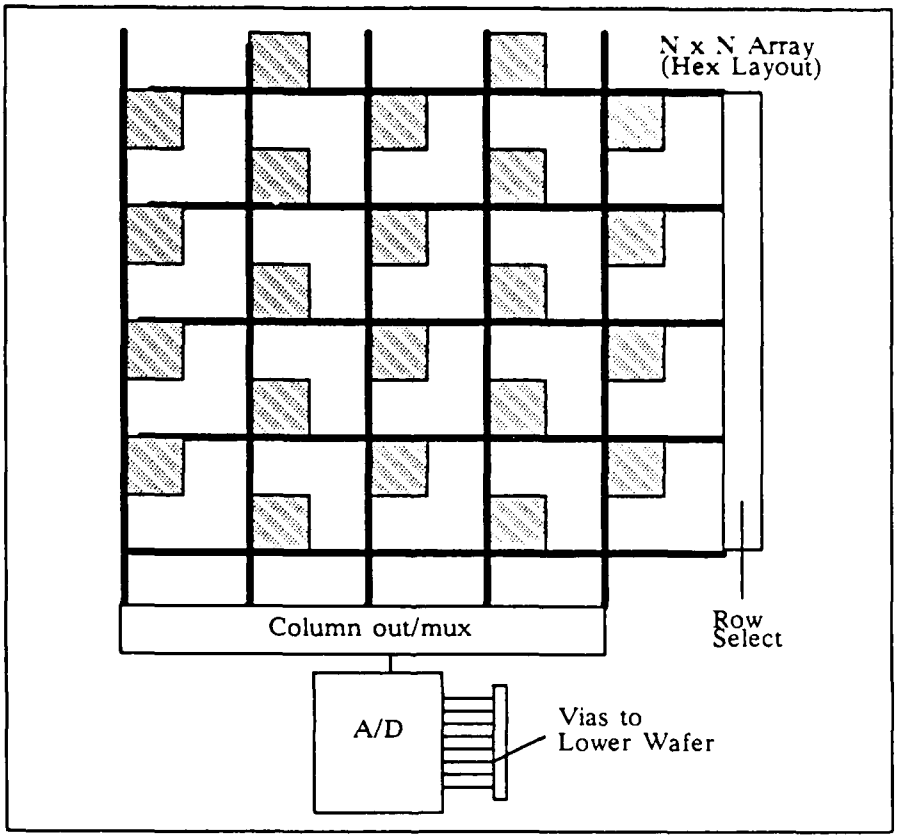


Figure 4-8: FPA I/O SCHEMATIC

Figure 4-9 is a CIF plot of a $15\mu m$ square photodiode. This was fabricated both with overglass and without for comparison purposes. The devices are not connected to pins but have probe pads for Vdd and output connections. Figure 4-10 is a CIF plot of a photodiode connected to an output driver. There is a variable load transistor to allow evaluation of a switching point. The device uses pin 24 for Vdd and pin 5 for GND. Pin 2 is the variable load voltage and pin 28 is the sensor output. Additionally, a photodiode similar to Figure 4-10 was fabricated but with a fixed 9 to 1 load transistor. This transistor uses Vdd and GND as above but pin 6 for the output.

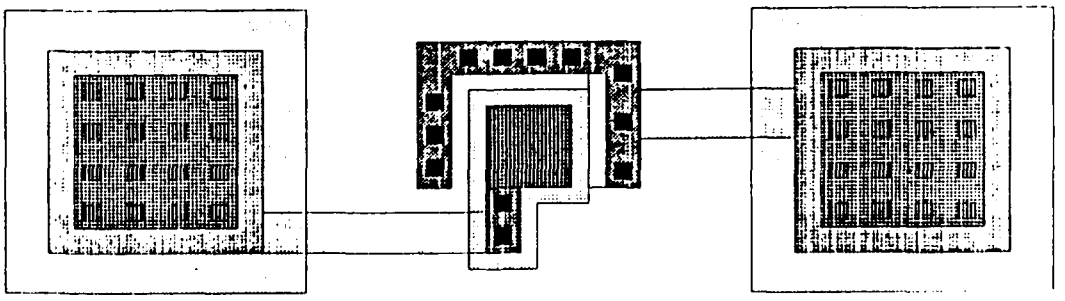


Figure 4-9: PHOTODIODE

In the monolithic architecture, a photodiode as above but larger ($30\mu m$ square) would take the place of the phototransistor in the previously discussed primitive cell. The output would be vertical through a via in the substrate. The primitive cell minus the pixel detector would be located directly below.

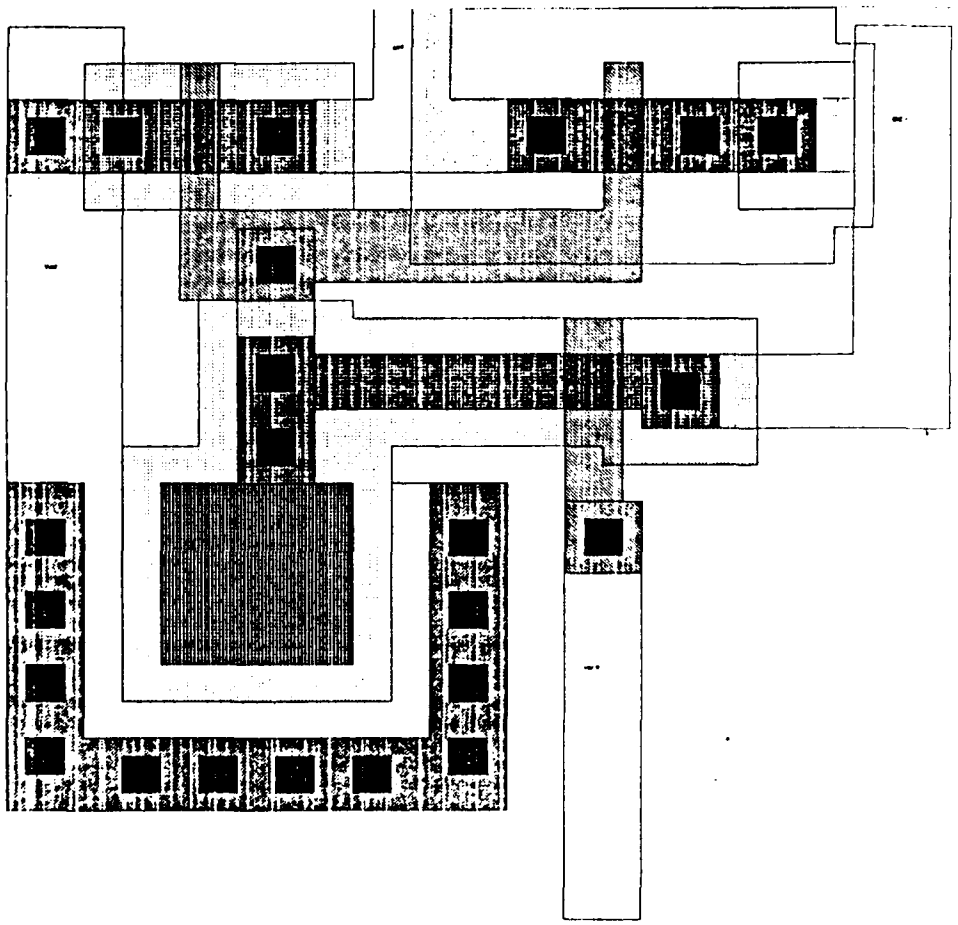


Figure 4-10: PHOTODIODE WITH DRIVER

4.4 3D Implementation

The details of how the 2D sensor array operates has been given. The 3D architecture allows signals to run in the vertical direction. Since the focal plane wafer is the top wafer of the stack it must get commands and provide output on the vertical bus. Figure 4-11 shows the basic theory of the 3D architecture.

4.4.1 3D Control Circuitry. These are the counterparts to the 2D circuits. The 3D architecture has control and wafer address lines running vertically through all the wafers. The bottom wafer in the stack (in this 3D implementation) is a special I/O wafer providing all external connections to the stack. In the monolithic design, the control lines terminate at the lower level of circuits and do not reach the top level where the photodiodes are. As explained in Appendix A, the number of control lines can equal the

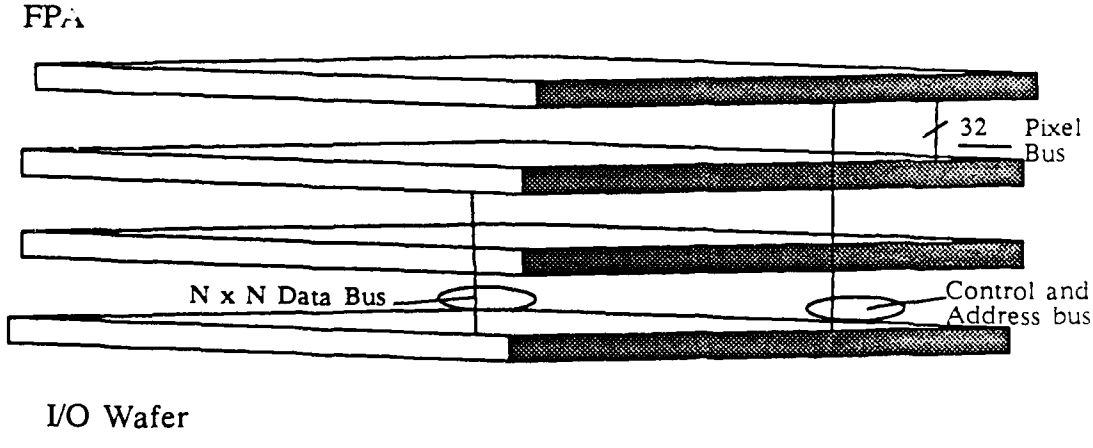


Figure 4-11: 3D ARCHITECTURE

number of rows or columns or their sum in the array because the control lines are actually fabricated by extending the array size. Therefore, a 512 x 512 3D computer can actually have 512, or even 1024 control lines if desired.

4.4.2 Input/Output. Every wafer except the FPA has a 512 x 512 data bus capability in addition to the control and addressing bus detailed above. The FPA wafer has the control and addressing bus but the data bus only consists of four 8 bit lines. These lines are separate from the other stack data bus. Stack I/O is provided by the I/O wafer at the bottom of the stack.

4.5 Summary

This chapter described the implementation of several devices used in the 3D electronic retina (any of the proposed architectures). These devices were design to facilitate testing and evaluation of the various functions in preparation of implementing a full scale model retina. Furthermore, photosensor devices were implemented that can be used in fabricating a monolithic 3D design.

The 3D routing architecture was also described. While the first full scale model may be implemented in 2D, the 3D routing constraints need to be anticipated to insure compatibility. This chapter provided the implementation details required to fabricate a fullscale electronic retina in any of the described architectures in Chapter 3.

CHAPTER 5

Results

5.1 Overview

The previous chapters described the mapping of the biological retina to an electro-optic design. System requirements were outlined and four different architectures were evaluated to determine the best overall in meeting those requirements. The key processing components were implemented and their functions explained. The material in this chapter concludes the research and discusses recommendations.

5.2 Results

The 3D architecture offers many benefits to the application of the electronic retina. The use of analog circuitry to perform the low level processing provides a means to speed up processing while reducing I/O requirements because of the data reduction. The combination of the 3D architecture with the analog low-level processing provides a system with the capability to provide real-time image processing in compact dimensions.

Qualitative information from other research efforts was used to a large extent. The photodetectors and processing cells implemented in this effort arrived in the eleventh hour and were not extensively tested. However, the testing accomplished thus far shows that the primitive cell is functional. The phototransistor current varies from the picoampere range to the microampere range with increasing light intensity. A commercial (Hamamatsu S1337-66BR photodiode) was used as an expedient to validate the test setup and estimate light intensity levels in lieu of a calibrated light source.

5.3 Conclusions

There is still much to be learned about the functioning of the biological retina. However, recent advances in the areas of 3D device fabrication and VLSI have provided the means to implement, at least to a limited degree, what is known about the function of the retina. Ideally, perfection of 3D circuit fabrication techniques will allow for devices to be arrayed vertically so the electronic retina will come even closer to the true retina architecture. Presently the 3D architecture available does allow a close approximation. Also as devices become more miniature, higher densities will be attainable.

It should be explicitly noted that the emulation of a retina in architecture and function is not done just for the sake of copying a biological system. Many years of evolution does not guarantee that a biological entity is in fact the best design because of the unique constraining factors of a biological organism and the associated complex environmental factors related to survival. However, as has been amply illustrated, the analog functions and the 3D architecture does in this case provide a sensible model providing superior image processing capabilities for many applications compared to contemporary designs.

The 3D approach is a viable and sensible approach to emulating the biological retina due to the simularity in architecture. The main constraint in implementing a total retina and harnessing the full capability of the architecture stems from not knowing how the retina functions.

5.4 Recommendations

This effort represents another step towards building a 3D electronic retina. Based on advanced concepts in neural networks and advances in VLSI and 3D architectures, the

3D electronic retina will provide a compact, real-time, low-level image processor.

The work should be continued by extensive testing of the implemented cells. A full scale array should be implemented. Refinements such as color capability and advanced motion discernablity can be applied.

APPENDIX A

Hughes 3D Computer Fabrication Technology

A.1 Introduction

This appendix outlines the basic methods and techniques used by Hughes to fabricate a 3-D computer. In addition to using conventional computer fabrication technologies, the Hughes 3-D computer fabrication required the development of new technologies to allow for the third axis of communication. These new technologies are feedthrough technology, microbridge technology, and 3-D assembly technology [Cor86].

A.2 Description

The concept of the Hughes 3-D computer involves the stacking of two or more wafers with a large number of z-axis channels between them (Figure A-1). Each wafer has an N x N array of circuits which are designed using normal wafer scale integration techniques (within the necessary design constraints to allow for the z-axis channels), with each element of the array associated with a unique z-axis channel.

In actuality, an additional M z-axis channels are provided for system control, system clocks, and power. If the array is extended by either one row or column, M = N, providing N system lines in a N x N + 1 array. The array could just as easily have been extended in both dimensions providing 2N system lines. For clarity and simplicity, all vertical channels are assumed to be included in the N x N array providing N^2 z-axis channels. Feedthrough technology provides the means by which the z-axis channels are formed, the microbridge technology provides the means by which the wafers are intercon-

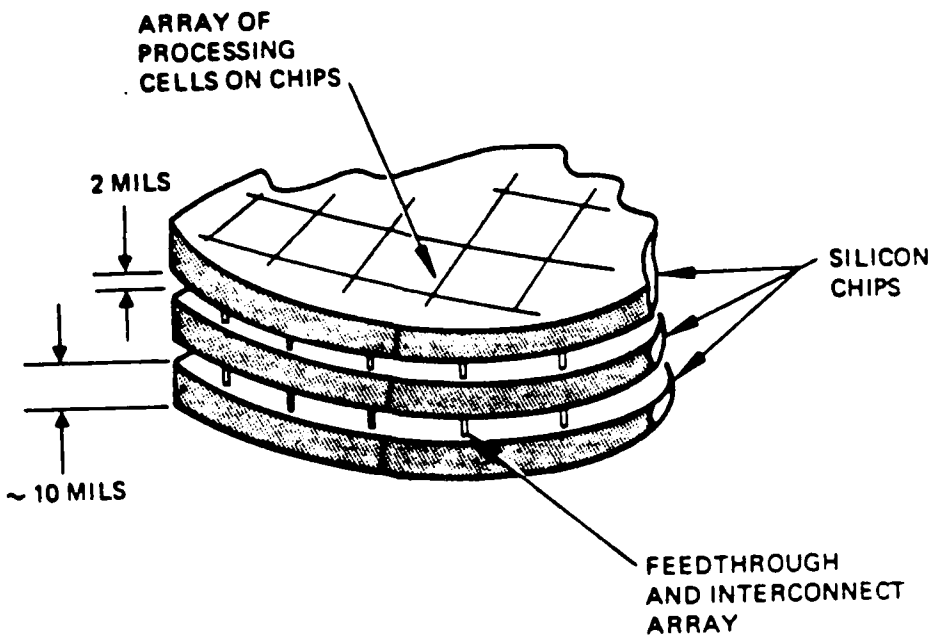


Figure A-1: STACKED WAFER CONCEPT
(from [Cor86])

nected vertically, and the Hughes 3-D assembly technology provides the means to hold the wafers together as an entity.

A.3 Feedthrough Technology

As previously stated, the feedthroughs provide channels through the wafer. These feedthroughs must be bi-directional and fast. Hughes has found the most suitable method of fabricating feedthroughs is thermomigration technology [Cor86]. The feedthrough resistance is typically between 25 and 30 ohms. The basic advantages of this technology include:

- Extremely short processing time. The migration rate of an aluminum droplet in silicon at 1150° C is ~1 mil/min. A typical wafer needs only a few tens of

minutes.

- Large aspect ratio. The lower limit of the area occupied by the bus line is determined mainly by the resistivity of the doped path. An aspect ratio of 10:1 is typical.
- Clean-room processing. Thermomigration is performed in an inert atmosphere with RF heating. The process conforms to the most stringent requirements of clean-room processing of silicon devices. The high yield of the process has already been demonstrated.
- Low-resistivity bus lines. The solid solubility of aluminum in silicon at 1100 ° C is 2 x 1019 cm-3. Assuming a bulk hole mobility of 100 cm2/V-s, the resulting resistivity will be ...

$$\rho = \frac{1}{qN_A\mu_\rho} = 3.1 \times 10^{-3} \ \ \Omega - cm.$$

- Integrity of silicon substrate. Other alternatives (such as etching holes through the silicon wafer and metallizing the hole walls) are complicated and would weaken the substrate [Cor86].

"Thermomigration is the phenomenon of a liquid zone (in the" form of a droplet, a sheet, or a rod) migrating in a solid along a thermal gradient" [Cor86]. Figure A-2 shows the migration of a liquid droplet through a solid. The forward edge of the droplet is warmer than the trailing edge forming a thermal gradient. Because the solubility of a solid in a liquid increases with temperature, the concentration of dissolved solid atoms is higher at the forward edge of the droplet than at the trailing edge, producing a concentration gradient of the dissolved solid through the droplet [Cor86]. "This concentration gradient, in turn, generates a diffusion" flux of dissolved solid atoms from the front to the rear interface [J in Figure A-2] of the droplet" [Cor86]. The presence of the diffusion flux causes the hot forward edge to dissolve more solid particles to replace the dissolved particles solidifying at the colder rear edge. Starting with an n-type silicon substrate, the dissolved aluminum atoms dope the redeposited silicon as p-type.

Figure A-3 illustrates how the feedthroughs are fabricated. The process starts with a silicon wafer about 5 to 6 mils thicker than standard. An approximately 10 mil thick

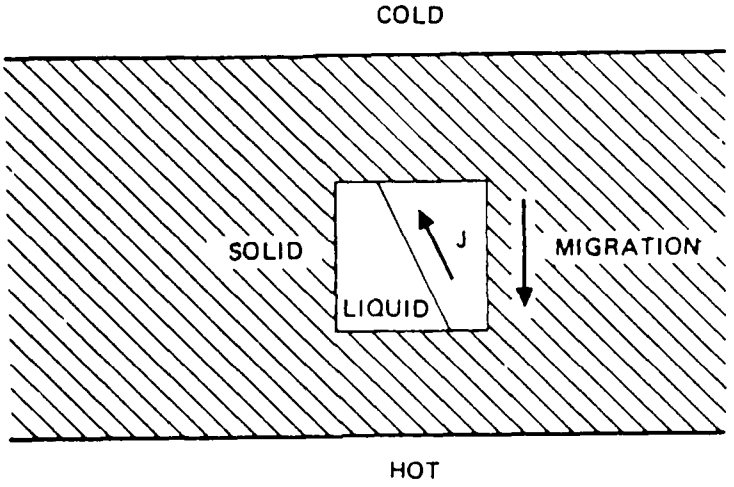


Figure A-2: THERMOMIGRATION
(from [Cor86])

layer of pure aluminum is deposited on this wafer. Using standard photolithography techniques and either a wet chemical or dry plasma etching process, portions of the aluminum layer are etched away leaving an array of aluminum dots on the wafer surface. Each dot represents the desired location of a channel [Cor86].

The wafer is then placed in a reactor specially designed and constructed to facilitate the thermomigration process. Upon completion of the thermomigration process, the wafer is lapped and polished on top and bottom which decreases the wafer thickness to standard size. The wafer is identical to standard wafers except for the presence of an N x N array of feedthroughs (channels through the wafer from top to bottom) [Cor86].

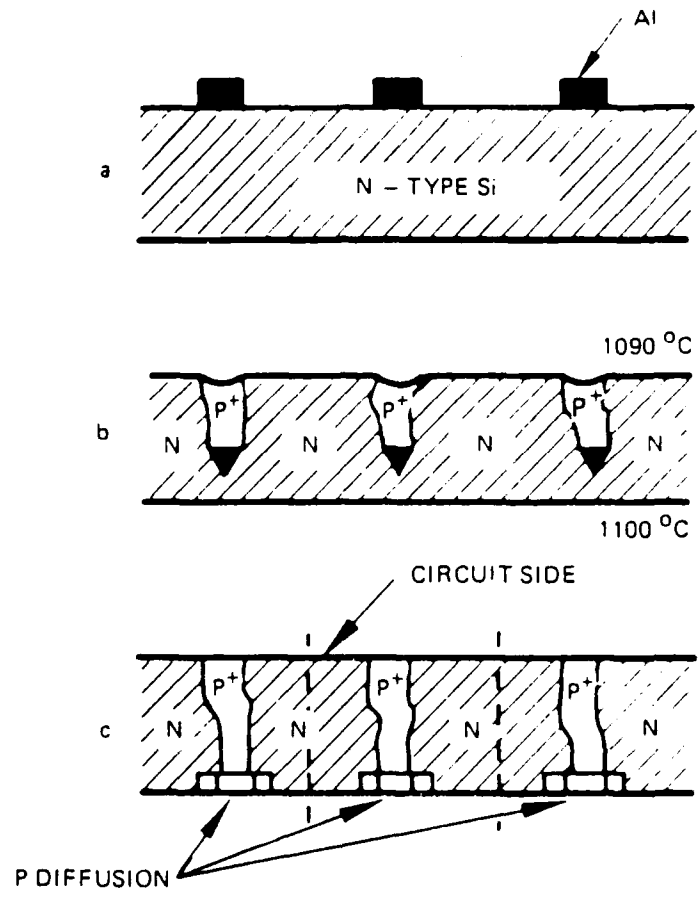


Figure A-3: FEEDTHROUGH FABRICATION
(from [Cor86])

A.4 Microbridge Technology

The microbridges provide the means to connect the feedthroughs of the stacked wafers allowing for an N x N array of vertical channels through the entire stack of wafers. Hughes developed this unique technology to provide fast wire-like performance, small size, and high reliability, while allowing for possible wafer warping. The

microbridge may be likened to a spring contact.

The microbridge is pictured in Figure A-4. Figure A-5 illustrates how the microbridges of adjacent wafers make contact at 90 degree angles. Because the springs make contact at right angles, this allows for a large possible misalignment of wafers when they are stacked. Additionally, the height of the microbridge allows for a significant amount of wafer warp while still providing a good contact. The spacers between wafers maintain wafer separation at 2 mils. Thus when the microbridges make contact they will flex approximately 16 microns ensuring a good contact.

The microbridge is fabricated by vacuum evaporation. Initially, a 10 micron thick spacer is evaporated or electroplated onto the substrate (Figure A-6a). Next, the spring contact is evaporated on top of the spacer (Figure A-6b). Also at this time, an outer coating of solder is vacuum deposited on the contact. When the spacer is etched away (Figure 6c) the microbridge is left. Upon assembly of a wafer stack, they are heated to the melting point of the solder providing a fused connection [Cor86].

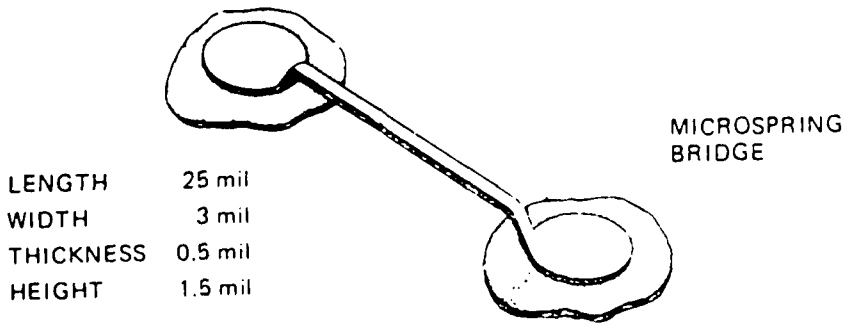


Figure A-4: A MICROBRIDGE CLOSEUP
(from [Cor86])

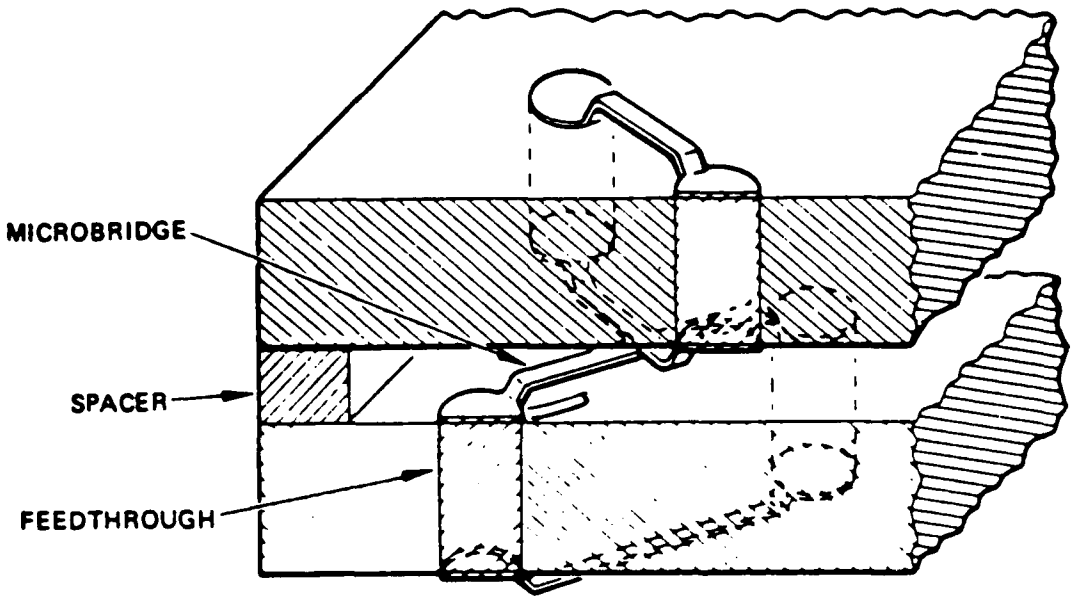


Figure A-5: WAFER INTERCONNECTION
(from [Cor86])

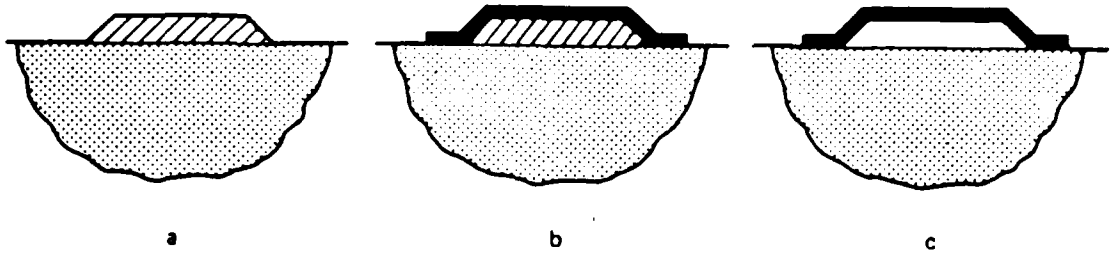


Figure A-6: MICROBRIDGE FABRICATION
(from [Cor86])

A.5 3-D Assembly Technology

The assembly of the 3-D computer provides for the following requirements:

- Wafers aligned with respect to one another with a precision better than the interconnect size
- Distance between wafers defined better than the interconnect compliance range
- Mechanical stability and rigidity adequate to prevent separation of or damage to the interconnects
- Thermal conductance sufficient to carry the heat generated by the stack circuits to the outside surface for dissipation
- Environmental protection
- Data I/O and power [Cor86].

Figure A-7 details the Hughes 3-D assembly. Two metal tabs are aligned and bonded to each wafer. The tabs have a precision hole on them. The wafers are aligned by sliding them onto the precision pins which are part of the assembly housing. After all the wafers are stacked and tested, they are heated to the melting point of the solder to fuse the microbridges together. After all the wafers have been stacked on the precision pins, the top plate is aligned on the precision pins and screwed down tightly. The top and

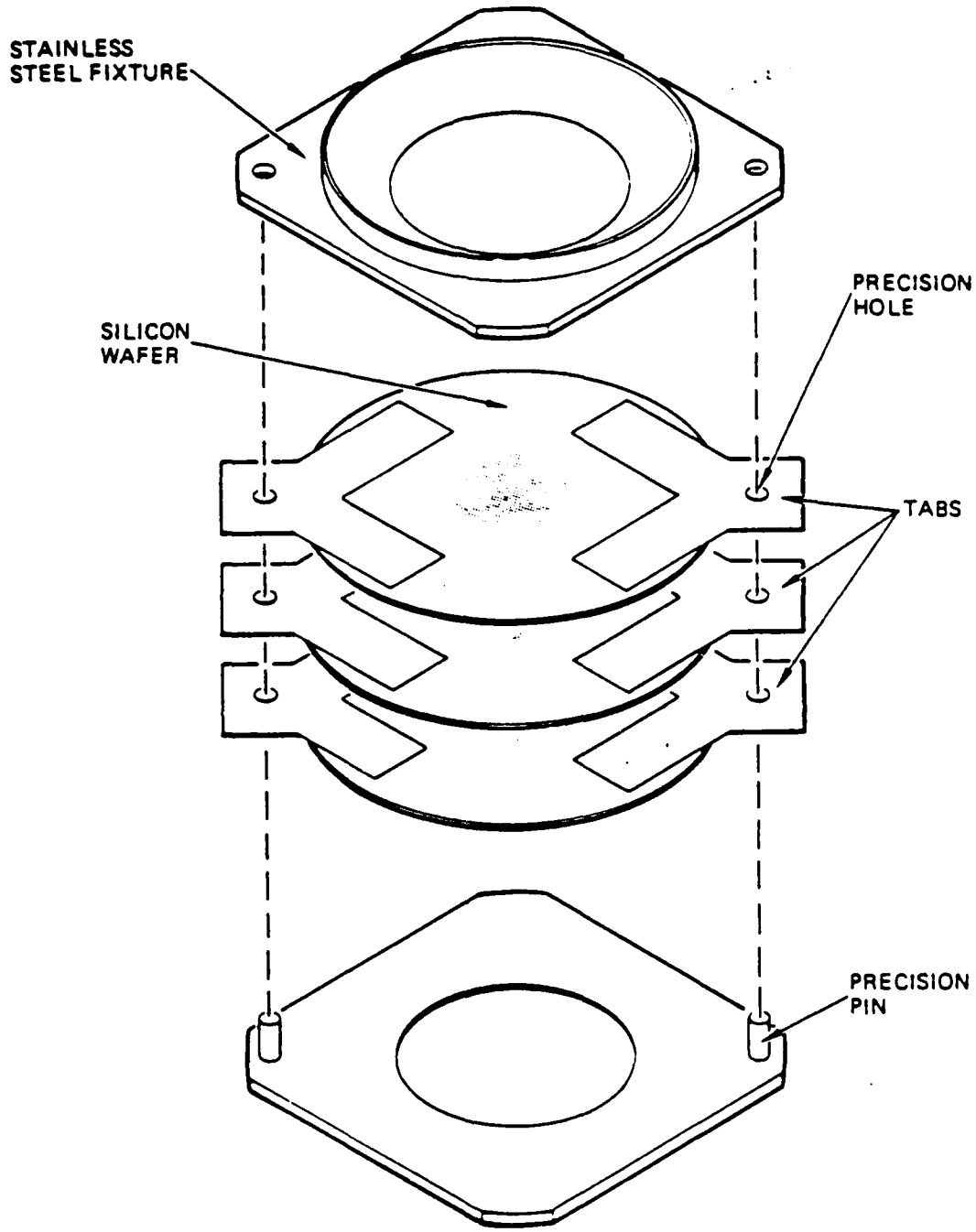


Figure A-7: 3D ASSEMBLY
(from [Cor86])

bottom plates made of stainless steel make an effective housing holding the wafer stack together [Cor86].

The stack I/O is accomplished through an edge connector using a replicator plane wafer. Specially designed I/O wafers can be fabricated to suit a given system requirement. Additionally, the stack controller is presently a separate microprocessor driven computer. Future designs will eventually incorporate an in-stack wafer scale system controller.

The above described apparatus has been thoroughly tested using 2 inch wafers. Due to the different thermal expansion coefficients of stainless steel and silicon, as well as possible heat dissipation problems, modifications to this procedure may be necessary for larger wafers [Cor86].

A.6 Summary

The basic 3-D computer fabrication technologies developed by Hughes allows for the fabrication of reliable and powerful 3-D computers. As these technologies are refined they will allow for larger wafer size and thus larger array sizes leading to more powerful computers.

APPENDIX B

Electromagnetic Spectrum

This appendix contains a figure showing the relationship between wavelength, frequency, and photon energy. It also provides an easy means of conversion between them.

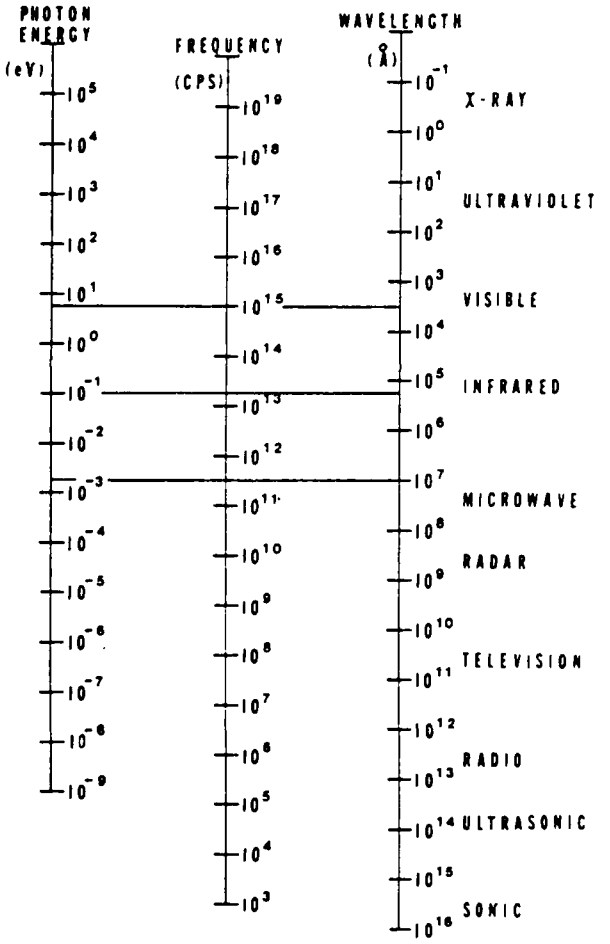


Figure B-1: ELECTROMAGNETIC SPECTRUM CHART
(from [Eli79])

APPENDIX C

Absorption Coefficients

This appendix contains a figure showing the absorption coefficients of various detector materials.

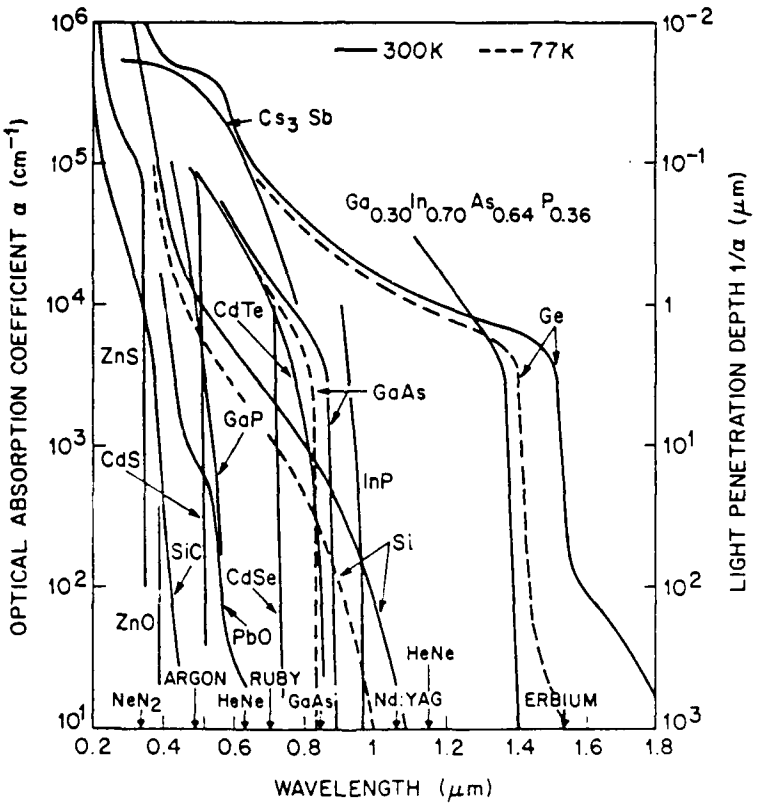


Figure C-1: ABSORPTION COEFFICIENTS
(from [Sze81])

Bibliography

[Bea82] Beatty, John C. and Kellogg S. Booth, editors, "Color Vision," in *Tutorial: Computer Graphics*, IEEE Computer Society Press, Silver Springs, Maryland (1982).

[Ber86] Bergman, L. A. and others, "Holographic Optical Interconnects for VLSI," *Optical Engineering* **25**(10) pp. 1109-1118 (October 1986).

[Bey80] Beynon, J. D. E. and D. R. Lamb, editors, *Charge-Coupled Devices and Their Applications*, McGraw-Hill, New York (1980).

[Cor70] Cornsweet, Tom N., *Visual Perception*, Academic Press, New York (1970).

[Cor86] Corporation, Hughes, "Development of a 3-D Processor," Technical Proposal 86M-0203/G1146, Hughes Research Laboratories (March 1986).

[Deg86] Degroat, Joseph and others, "Space Surveillance Signal Processing," A White Paper Proposal, Air Force Institute of Technology, Wright-Patterson AFB OH (27 October 1986).

[Eli79] Elion, Glenn R. and Herbert A. Elion, *Electro-Optics Handbook*, Marcel Dekker, New York (1979).

[Fos87] Fossum, Eric R., "Charge-Coupled Computing for Focal Plane Image Preprocessing," *Optical Engineering* **26**(9) pp. 916-922 (September 1987).

[Gri84] Grinberg, Jan and others, "3-D Computing Structures for High Throughput Information Processing," in *VLSI Signal Processing*, ed. Leah Jamieson,IEEE Press, New York (1984).

[Gri4A] Grinberg, Jan and others, "A Cellular VLSI Architecture," *Computer* **17** pp. 69-81 (January 1984A).

[Hec87] Hecht, Eugene, *Optics*, Addison-Wesley, Reading, Massachusetts (1987).

[Jan87] Janesick, James R., "Charge-Coupled-Device Characterization, Modeling and Application," *Optical Engineering* **26**(8) p. 685 (August 1987).

[Kab66] Kabrisky, Matthew, *A Proposed Model for Visual Information Processing in the Human Brain*, University of Illinois Press, Urbana and London (1966).

[Kat85] Kataoka, Shoei, "An Attempt Towards an Artificial Retina: 3-D IC Technology for an Intelligent Image Sensor," pp. 440-442 in *1985 International Conference on Solid-State Sensors and Actuators*, IEEE, New York (1985).

[Key80] Keyes, R. J., "Introduction," in *Optical and Infrared Detectors*, ed. R. J. Keyes,Springer-Verlag, New York (1980).

[Lyo81] Lyon, Richard F., "The Optical Mouse, and an Architectural Methodology for Smart Digital Sensors," in *VLSI Systems and Computations*, ed. H. T. Kung and others,Computer Science Press, Rockville, Maryland (1981).

[Mas86] Masland, Richard H., "The Functional Architecture of the Retina," *Scientific American* **255** pp. 102-111 (December 1986).

[Mea87] Mead, Carver A., "Neural Hardware for Vision," *Engineering & Science* L pp. 2-7 (June 1987).

[Mic80] Michon, G. J. and H. K. Burke, "CID Image Sensing," in *Charge Coupled Devices*, ed. D. F. Barbe,Springer-Verlag, New York (1980).

[Mon82] Montalvo, Fanya S., "Special Session on Vision," in *Tutorial: Computer Graphics*, IEEE Computer Society Press, Silver Springs, Maryland (1982).

[Pog87] Poggio, Thomas and Christof Koch, "Synapses That Compute Motion," *Scientific American* J pp. 46-52 (May 1987).

[Rat61] Ratliff, Floyd, "Inhibitory Interaction and the Detection and Enhancement of Contours," in *Sensory Communication*, ed. Walter A. Rosenblith,John Wiley & Sons, New York (1961).

[Ros79] Ross, Douglas A., *Optoelectronic Devices and Optical Imaging Techniques,* The MacMillan Press, London (1979).

[Siv87] Sivilotti, Massimo A. and others, "Real-Time Visual Computations Using Analog CMOS Processing Arrays," Research Report, California Institute of Technology, Pasedena, California (1987).

[Sze81] Sze, S. M., *Physics of Semiconductor Devices,* John Wiley & Sons, New York (1981).

[Weh85] Wehner, William W., "Modular Real Time Architectures for Image Understanding Applications," *Architectures and Algorithms for Digital Image Processing II* **534** pp. 12-21 (January 1985).

[Wer73] Werblin, Frank S., "The Control of Sensitivity in the Retina," *Scientific American* **228** pp. 71-79 (January 1973).

[Wes85] Weste, Neil and Kamran Eshraghian, *Principles of CMOS VLSI Design,* Addison-Wesley, Reading, Massachusetts (1985).

[Wya87] Wyatt, Clair L., *Radiometric System Design,* MacMillan Publishing Company, London (1987).

Vita

Captain Gray L. Salada was born 23 March 1954 in Heidelberg, West Germany. After graduating from Kecoughtan Senior High School, Hampton, Virginia in 1972, he joined the army. He attended the West Point Preparatory School at Fort Belvoir, Virginia and subsequently the United States Military Academy at West Point where, on June 8, 1977, he received a Bachelor of Science Degree in Engineering, and a commission in the U. S. Army Signal Corps.

After attending the Signal Officer basic course, he was assigned to the Republic of Korea as a platoon leader in the 304th Signal Battalion in 1977. He was reassigned to the 82d Airborne Division at Fort Bragg, North Carolina where he served as Communications officer in the 1/320th Field Artillery (Abn) from 1979 to 1980, and as the Communications officer in the 2/504th Airborne Infantry from 1981 to 1982. In 1982 Captain Salada attended the Signal Officer advanced course at Fort Gordon, Georgia and the Teleprocessing Operations Course at the Air Force Institute of Technology, Wright Patterson AFB, Ohio.

In 1983 Captain Salada was reassigned to Heidelberg, West Germany where he served on the Headquarters U.S. Army Europe Staff prior to commanding the 178th Signal Company from 1985 to 1986 . During his three year tour in Germany, Captain Salada earned a Master of Science Degree in Systems Management from the University of Southern California. Upon his return to the states in 1986, he attended the Combined Arms Staff School, Fort Leavenworth, Kansas. In June 1986 Captain Salada entered the School of Engineering, Air Force Institute of Technology.

Permanent Address: P.O. Box 476

Melrose, Florida 32666

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release; Distribution Unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/GCS/ENG/87D-23	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION School of Engineering 6b. OFFICE SYMBOL (If applicable) AFIT/ENG	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Air Force Institute of Technology Wright-Patterson AFB, OH 45433	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Ofc of Scientific Research 8b. OFFICE SYMBOL (If applicable) AFOSR/XP	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Bolling AFB, Washington, D.C. 20332	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

A THREE DIMENSIONAL ELECTRONIC RETINA ARCHITECHTURE, UNCLASSIFIED

12. PERSONAL AUTHOR(S)

Salada, Gray L., CPT, U.S. Army

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
MS Thesis	FROM _____ TO _____	1987 December 7	93

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Eye, Retina, Phototransistors, Semiconductor, Photodiodes, Image Processing
12	09		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

THESIS ADVISOR: Richard W. Linderman, Capt, USAF
Assistant Professor of Electrical Engineering

Approved for public release IAW AFR 190-1.
LYNN E. WOLAVER 31 Dec 87
Dean for Research and Professional Development
Air Force Institute of Technology (AU)
Wright-Patterson AFB OH 45433

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED	
22a NAME OF RESPONSIBLE INDIVIDUAL Richard W. Linderman, Capt, USAF	22b TELEPHONE (Include Area Code) 513-255-3576	22c OFFICE SYMBOL AFIT/ENG

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

The biological retina is a compact, real-time image processor. It performs analog computations in parallel in depth. Such processing capabilities have so far been unattainable in comparably sized man-made image processors. This thesis investigates using VLSI technology and a three dimensional architectural approach in designing an electronic retina.

This effort involves the comparison of alternative 3D designs and the implementation of VLSI circuitry capable of low-level image processing. The 3D architectures compared include a model based on the Hughes 3D computer, a model using a monolithic wafer with multiple circuit layers, and a model using stacked wafers with fiber optic interconnects. A conventional 2D architecture is presented for comparison/reference.

Circuits for photo detection, and spatial and temporal processing were implemented. It was shown that using VLSI analog circuits and a 3D architecture it is feasible to fabricate a real-time image processor in a compact package.